

SEARCH REQUEST FORM

Requestor's

Name: _____

Serial

Number: _____

Date: _____

Phone: _____

Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 1/29

Searcher: D. Schreiber 308-4292

Terminal time: 6

Elapsed time: 6

CPU time: _____

Total time: _____

Number of Searches: _____

Number of Databases: _____

Search Site

☐ STIC

☒ CM-1 6A03

☐ Pre-S

Type of Search

☐ N.A. Sequence

☒ A.A. Sequence

☐ Structure

☐ Bibliographic

Vendors

☐ IG

☐ STN

☐ Dialog

☐ APS

☐ Geninfo

☐ SDC

☐ DARC/Questel

☒ Other Computer

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 06:44:44 : Search time 31.5439 Seconds
(without alignments)
122.505 Million cell updates/sec

Title: US-09-627-165D-19

Perfect score: 153
Sequence: 1 YERKRLRYHTGTGDPYFKFTLLADQHS 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121	79.1	254	22 AAB47091	A-chain isoform fo
2	114	74.5	28	15 AAR44926	Galactoside-specif
3	114	74.5	254	22 AAB47090	A-chain isoform fo
4	114	74.5	256	20 AAY25981	Mistletoe lectin A
5	114	74.5	256	20 AAY25984	Mistletoe lectin A
6	114	74.5	551	23 ABP79450	Galactose-recognt
7	114	74.5	564	18 AAR10021	Prepro mistletoe I
8	114	74.5	564	20 AAW50127	Mistletoe lectin p
9	110	71.9	252	19 AAW64659	Mistletoe tMLA pro
10	110	71.9	252	19 AAW64661	Mistletoe tMLA var

11	110	71.9	253	18 AAW10022	Prepro mistletoe I
12	110	71.9	253	20 AAW90125	Mistletoe ML A-cha
13	110	71.9	254	20 AAY25980	Mistletoe lectin A
14	110	71.9	254	20 AAY25983	Mistletoe lectin A
15	110	71.9	531	20 AAY25979	Mistletoe lectin I
16	110	71.9	532	20 AAY25982	Mistletoe lectin I
17	109	71.2	256	22 AAB47092	A-chain isoform fo
18	107	69.9	255	20 AAY25971	Mistletoe lectin A
19	107	69.9	255	20 AAY25974	Mistletoe lectin A
20	107	69.9	255	20 AAY25977	Mistletoe lectin A
21	107	69.9	533	20 AAY25970	Mistletoe lectin p
22	107	69.9	533	20 AAY25973	Mistletoe lectin p
23	107	69.9	533	20 AAY25976	Mistletoe lectin p
24	88	57.5	20	15 AAR44925	Galactoside-specif
25	48.5	31.7	56	23 ABP02301	Human ORFX protein
26	47.5	31.0	385	22 AAG17516	Human ORFX protein
27	47	30.7	318	22 AAG71841	Human olfactory re
28	47	30.7	318	22 AAG72444	Human OR-like poly
29	47	30.7	2536	22 AAB63874	Drosophila melanog
30	46	30.1	9	20 AAW90128	Mistletoe lectin M
31	45.5	29.7	1326	22 AAB59965	Drosophila melanog
32	45	29.4	267	22 AAB50827	Human shear stress
33	45	29.4	758	22 ABB58337	Drosophila melanog
34	45	29.4	1120	22 ABB507425	Novel human diagno
35	45	29.4	1267	22 ABB522790	Novel human diagno
36	45	29.4	1560	21 AAY67203	Novel human diagno
37	45	29.4	1561	21 AAY72202	Novel human diagno
38	45	29.4	1562	21 AAB18639	S. venezuelae synta
39	45	29.4	1562	21 AAY72194	Amino acid sequenc
40	45	29.4	5035	13 AAR25450	S. venezuelae macr
41	45	29.4	12199	21 AAY77180	Mt mutant porcine
42	44.5	29.1	374	23 ABB59036	S. venezuelae plk
43	44.5	29.1	378	20 AAY34744	Chlamydia pneumoni
44	44.5	29.1	702	23 AAB04353	Protein SCNM1A dif
45	44	28.8	72	22 ABB28517	Peptide #1168 enco

ALIGNMENTS

RESULT 1

AAB47091
ID AAB47091 standard. Protein: 254 AA.

AC AAB47091:

DT 16-MAY-2001 (first entry)

XX A-chain isoform for biosynthesis of a Korean mistletoe lectin #2.

DE Isoform: A-chain; B-chain: biosynthesis: lectin; Korean mistletoe;

KW KML: tumour; KM-110; KML-C; KMBB; KML-IIU; KML-III;

KV heparin binding protein.

XX Viscum album coloratum.

OS Key Location/Qualifiers

XX Key Misc-difference 240 /note= "Encoded by YTA"

XX EPI074560-AZ.

PD 07-FEB-2001.

XX 27-JUL-2000: 2000EP-0402168.

XX 27-JUL-1999: 99KR-0030638.

XX (MIST-) MISTLE BIOTECH CO LTD.

XX Kim J, Song S, Suh B, Lee K, Doo M, Kwak J, Song B, Yoon T;

PI Kang T, Park C;

XX

DR WPI: 2001-171044/18.
DR N-PSDB: AAC85473.

XX Novel lectin proteins isolated from Korean mistletoe, useful for
PT enhancing immunity and effectuating anti-tumoral activity -
XX

PS Claim 4; Page 26-27; 62pp; English.

XX The sequences given in AAB47090-92 are isoforms of an A-chain gene
CC involved in biosynthesis of lectins isolated from Korean mistletoe.
CC Korean mistletoe lectins (KML) are useful for enhancing immunity and
CC for treating tumours. The KML's are isolated from a protein fraction
CC derived from the leaves, stems and fruits of Korean mistletoe, which
CC is designated KM-110. One of the isolates, KML-C was shown to be
CC effective against colon 26-M3.1 carcinoma and L5178Y-M25 lymphoma.
XX

SO Sequence 254 AA;

Query Match 79.1%; Score 121; DB 22; Length 254;
Best Local Similarity 82.8%; Pred. No. 7.8e-11;

Matches 24; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 YERKRLRVHTQTGDTGDFKFTLLADQS 29
111 :|||||
1 YERLRLRVHTQTGDTGDFKFTLLRDHVS 29

RESULT 2

ID AAB44926
AAB44926 standard; peptide; 28 AA.

AC AAB44926;

DT 04-JUL-1994 (first entry)

DE Galactoside-specific lectin ML-I A2 chain N-terminal.

KW Mistletoe; immunomodulatory; cancer; therapy; treatment; adjuvant;

KW supportive; palliative; toxic.

OS Viscum album.

XX DE4221836-A.

PN 05-JAN-1994.

XX 03-JUL-1992; 92DE-4221836.

XX 03-JUL-1992; 92DE-4221836.

XX 03-JUL-1992; 92DE-4221836.

PA (BARD/) BARDOSI A.
PA (GABI/) GABIUS HJ.

DR WPI: 1994-008729/02.

PT New mistletoe lectin - with immunomodulatory activity, useful for
PT cancer adjuvant therapy

PS Claim 3; Page 5; 8pp; German.

XX The sequence is that of the N-terminal of the toxic A2 chain
CC of the mistletoe galactoside-specific lectin (ML-I). ML-I may
CC be used for adjuvant therapy in the surgical, chemotherapeutic
CC and radiotherapeutic treatment of cancer.
XX

SO Sequence 28 AA;

Query Match 74.5%; Score 114; DB 15; Length 28;
Best Local Similarity 80.8%; Pred. No. 6.6e-11;

Matches 21; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKRLRVHTQTGDTGDFKFTLLAD 26
111 :|||||

DB 1 YERLRLRVHTQTGDTGDFKFTLLRD 26

RESULT 3

ID AAB47090
AAB47090 standard; protein; 254 AA.

AC AAB47090;

DT 16-MAY-2001 (first entry)

DE A-chain isoform for biosynthesis of a Korean mistletoe lectin #1.

KW Isoform: A-chain; B-chain; biosynthesis; lectin; Korean mistletoe;
KW KML; tumour; KM-110; KML-C; KMHP; KML-110; KML-11L;
KW heparin binding protein.

OS Viscum album coloratum.

PN EP1074560-A2.

PD 07-FEB-2001.

PF 27-JUL-2000; 2000EP-0402168.

PR 27-JUL-1999; 99KR-0030638.

PA (MIST-) MISTLE BIOTECH CO LTD.

PI Kim J, Song S, Suh B, Lee K, Doo M, Kwak J, Song B, Yoon T;
PI Kang T, Park C.

DR WPI: 2001-171044/18.

DR N-PSDB: AAC85472.

PT Novel lectin proteins isolated from Korean mistletoe, useful for
PT enhancing immunity and effectuating anti-tumoral activity -

PS Claim 2; Page 25-26; 62pp; English.

XX The sequences given in AAB47090-92 are isoforms of an A-chain gene
CC involved in biosynthesis of lectins isolated from Korean mistletoe.
CC Korean mistletoe lectins (KML) are useful for enhancing immunity and
CC for treating tumours. The KML's are isolated from a protein fraction
CC derived from the leaves, stems and fruits of Korean mistletoe, which
CC is designated KM-110. One of the isolates, KML-C was shown to be
CC effective against colon 26-M3.1 carcinoma and L5178Y-M25 lymphoma.
XX

SO Sequence 254 AA;

Query Match 74.5%; Score 114; DB 22; Length 254;
Best Local Similarity 80.8%; Pred. No. 9.9e-10;

Matches 21; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKRLRVHTQTGDTGDFKFTLLAD 26
111 :|||||
1 YERLRLRVHTQTGDTGDFKFTLLRD 26

RESULT 4

ID AAY25981
AAY25981 standard; protein; 256 AA.

AC AAY25981;

DT 18-OCT-1999 (first entry)

DE Mistletoe lectin A2 protein fragment.

KW Mistletoe; lectin; antitumour; immunostimulant; A-chain; MLA; immunity;
KW ribozyme 28S subunit; non-cytotoxic; T-cell activation; immune response;
KW lymphokine-producing macrophage; uncontrolled cell growth; treatment;
KW cancer; cytotoxicity; antigen; isoform; lectin A2.


```

XX      Viscum album.
OS      DE19804210-A1.
XX      12-AUG-1999.
XX      03-FEB-1998; 98DE-1004210.
XX      03-FEB-1998; 98DE-1004210.
XX      03-FEB-1998; 98DE-1004210.
XX      (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX      Morris P, Stiefel T, Voelter W, Welters P;
XX      WPI: 1999-445335/38.
XX      N-PSDB: AA209105.
XX      Preparation of mistletoe lectins in heterologous systems,
XX      particularly for use as anticancer agents and immunostimulants
XX      Claim 8; Fig 3B; 78pp; German.
XX      This invention describes a novel mistletoe lectin (I) and its fragments
XX      which have antitumour and immunostimulatory activity. The A-chain (MLA)
XX      of the mistletoe lectin binds to, and inactivates, the 28S subunit of
XX      ribosomes. Non-cytotoxic forms of (I) activate T-cell and
XX      lymphokine-producing macrophages, so stimulate immunity. (I) and its
XX      fragments are used to treat uncontrolled cell growth (particularly
XX      cancers) and if they lack cytotoxicity, to increase the strength of the
XX      immune response, particularly to a co-administered antigen
XX      (tumour-associated, bacterial or viral). The method allows production of
XX      mistletoe lectin, and its individual chains, in many different isoforms
XX      and on a large scale, at any time of the year. Recombinant products are
XX      free from toxins present in natural mistletoe extracts. This sequence
XX      represents a fragment of the mistletoe lectin A2 protein.
XX
SQ      Sequence 256 AA:
Query Match 74.5%; Score 114; DB 20; Length 256;
Best Local Similarity 80.8%; Pred. No. 1e-09;
Matches 21; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY      1 YEREKLRVTHQTGDDQYFKFRTLLAD 26
      111 :|||||:|||||:|||||
Db      1 YERLRVTHQTGDDYFRFTLLRD 26
RESULT 5
AAY25984
ID      AAY25984 standard; Protein: 256 AA.
XX
AC      AAY25984;
XX
DT      18-OCT-1999 (first entry)
XX
DE      Mistletoe lectin A2 (variant) protein fragment.
XX
KW      Mistletoe; lectin; antitumour; immunostimulant; A-chain; MLA; immunity;
KW      ribosome 28S subunit; non-cytotoxic; T-cell activation; immune response;
KW      lymphokine-producing macrophage; uncontrolled cell growth; treatment;
KW      cancer; cytotoxicity; antigen; isoform; lectin A2.
XX
OS      Viscum album.
OS
XX
PN      DE19804210-A1.
XX
PD      12-AUG-1999.
XX
PF      03-FEB-1998; 98DE-1004210.
XX
PR      03-FEB-1998; 98DE-1004210.
XX

```

```

PA      (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX      Morris P, Stiefel T, Voelter W, Welters P;
XX      WPI: 1999-445335/38.
XX      N-PSDB: AA209108.
XX      Preparation of mistletoe lectins in heterologous systems,
XX      particularly for use as anticancer agents and immunostimulants
XX      Disclosure: Fig 6B; 78pp; German.
XX      This invention describes a novel mistletoe lectin (I) and its fragments
XX      which have antitumour and immunostimulatory activity. The A-chain (MLA)
XX      of the mistletoe lectin binds to, and inactivates, the 28S subunit of
XX      ribosomes. Non-cytotoxic forms of (I) activate T-cell and
XX      lymphokine-producing macrophages, so stimulate immunity. (I) and its
XX      fragments are used to treat uncontrolled cell growth (particularly
XX      cancers) and if they lack cytotoxicity, to increase the strength of the
XX      immune response, particularly to a co-administered antigen
XX      (tumour-associated, bacterial or viral). The method allows production of
XX      mistletoe lectin, and its individual chains, in many different isoforms
XX      and on a large scale, at any time of the year. Recombinant products are
XX      free from toxins present in natural mistletoe extracts. This sequence
XX      represents a fragment of a mistletoe lectin A2 protein variant.
XX
SQ      Sequence 256 AA:
Query Match 74.5%; Score 114; DB 20; Length 256;
Best Local Similarity 80.8%; Pred. No. 1e-09;
Matches 21; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY      1 YEREKLRVTHQTGDDQYFKFRTLLAD 26
      111 :|||||:|||||:|||||
Db      1 YERLRVTHQTGDDYFRFTLLRD 26
RESULT 6
ABB79450
ID      ABB79450 standard; Protein: 551 AA.
XX
AC      ABB79450;
XX
DT      08-JUL-2002 (first entry)
XX
DE      Galactose-recognising mistletoe lectin.
XX
KW      Mistletoe; galactose-recognising mistletoe lectin; MLIII.
XX
OS      Viscum album.
XX
FH      Key Location/Qualifiers
FH      MISC-difference 223 /note= "Encoded by ATG"
FH      MISC-difference 251 /note= "Encoded by TTT"
FH      MISC-difference 344 /note= "Encoded by TCG"
FH      MISC-difference 380 /note= "Encoded by GCC"
FH      MISC-difference 448 /note= "Encoded by GTG"
XX
DE10044027-A1.
XX
PD      14-MAR-2002.
XX
PF      06-SEP-2000; 2000DE-1044027.
XX
PR      06-SEP-2000; 2000DE-1044027.
XX
PA      (VISC-) VISCUM AG.
XX

```

PI Kieff S;
 XX
 DR WPI: 2002-316737/36.
 DR N-PSDB; ABL56947.
 XX
 PT New nucleic acid encoding preproprotein of mistletoe lectin, useful as
 PT diagnostic and therapeutic agents, also encodes polypeptide -
 XX
 PS Claim 1; Fig 1; 6pp; German.
 XX
 CC The invention relates to a nucleic acid molecule (ABL56947) that encodes
 CC a preprotein (ABP79450) which, after maturation, has the biological
 CC activity of the galactose-recognising mistletoe lectin (MLIII). The MLIII
 CC encoding nucleic acid molecule, primers specific to it or complements of
 CC it, and encoded (oligomeric) polypeptides are useful as diagnostic and
 CC therapeutic agents.
 XX
 SQ Sequence 551 AA;
 Query Match 74.5%; Score 114; DB 23; Length 551;
 Best Local Similarity 80.8%; Pred. No. 2.6e-09;
 Matches 21; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 YEREKLRVTHQTGDDYEFKFTLLAD 26
 ||| :|||||||:|||||||
 Db 34 YERLRRLRVTHQTGDEYFRFTLLRD 59
 RESULT 7
 AAM10021
 ID AAM10021 standard; Protein; 564 AA.
 XX
 AC AAM10021;
 XX
 DT 18-DEC-1997 (first entry)
 XX
 DE Prepro mistletoe lectin.
 XX
 KM Mistletoe; lectin; cytotoxic; A chain; B chain; dimer.
 XX
 OS Viscum album.
 XX
 PN EP751221-A1.
 XX
 PD 02-JAN-1997.
 XX
 PF 26-JUN-1995; 95EP-0109949.
 XX
 PR 26-JUN-1995; 95EP-0109949.
 XX
 PA (MADU) MADAU5 KOELN AG.
 XX
 PI Baur A, Eck J, Lentzen H, Zinke H;
 XX
 DR WPI: 1997-054678/06.
 DR N-PSDB; AAT70473.
 XX
 PT Nucleic acid encoding prepro form of mistletoe lectin - for
 PT therapeutic or diagnostic use
 XX
 PS Claim 12; Fig 4c; 30pp; German.
 XX
 CC Mistletoe lectin is a cytotoxic agent that has been used for tumour
 CC therapy. It can be used in immunotoxins and medicaments. Nucleic
 CC acid fragments can be used in diagnostic methods. Mistletoe lectin (
 CC AAT70473) comprises an A chain (AAT70474) and a B chain (AAT70475).
 XX
 SQ Sequence 564 AA;
 Query Match 74.5%; Score 114; DB 18; Length 564;
 Best Local Similarity 80.8%; Pred. No. 2.6e-09;
 Matches 21; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 YEREKLRVTHQTGDDYEFKFTLLAD 26
 ||| :|||||||:|||||||
 Db 34 YERLRRLRVTHQTGDEYFRFTLLRD 59
 RESULT 8
 AAM90127
 ID AAM90127 standard; Protein; 564 AA.
 XX
 AC AAM90127;
 XX
 DT 30-APR-1999 (first entry)
 XX
 DE Mistletoe lectin prepro-protein.
 XX
 KM ML; mistletoe; lectin; ML; transgenic plant; glycosylation;
 KM dimer; immunotoxin; large-scale production; diagnosis; therapeutic;
 KM cancer.
 XX
 OS Viscum album.
 XX
 PN EP84388-A1.
 XX
 PD 16-DEC-1998.
 XX
 PF 26-JUN-1995; 95EP-0109949.
 XX
 PR 26-JUN-1995; 95EP-0109949.
 XX
 PR 26-JUN-1995; 98EP-0105660.
 XX
 PA (MADU) MADAU5 KOELN AG.
 XX
 PI Baur A, Eck J, Lentzen H, Zinke H;
 XX
 DR WPI: 1999-026582/03.
 DR N-PSDB; AAV74182.
 XX
 PT New transgenic plant expressing mistletoe lectin - useful for
 PT producing recombinant lectin in e.g. cancer diagnosis and therapy
 XX
 PS Claim 1a; Fig 4c; 30pp; German.
 XX
 CC This invention describes a novel transgenic plant transformed with a
 CC vector capable of encoding a mistletoe (Viscum album) lectin
 CC preprotein or a biologically active fragment. The specification
 CC also describes a polypeptide produced by a plant where the polypeptide
 CC exhibits at least one enzymatic modification other than the glycosylation
 CC that occurs in Viscum album or the polypeptide is a fusion protein, a
 CC mistletoe lectin polypeptide dimer and an immunotoxin comprising the
 CC polypeptide or the polypeptide dimer. The plants are used for large-scale
 CC production of mistletoe lectin for diagnostic or therapeutic purposes
 CC (e.g. in cancer therapy). This sequence represents the mistletoe lectin
 CC used in the method of the invention.
 XX
 SQ Sequence 564 AA;
 Query Match 74.5%; Score 114; DB 20; Length 564;
 Best Local Similarity 80.8%; Pred. No. 2.6e-09;
 Matches 21; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 YEREKLRVTHQTGDDYEFKFTLLAD 26
 ||| :|||||||:|||||||
 Db 34 YERLRRLRVTHQTGDEYFRFTLLRD 59
 RESULT 9
 AAM64659
 ID AAM64659 standard; Protein; 252 AA.
 XX
 AC AAM64659;
 XX
 DT 23-OCT-1998 (first entry)
 XX

```

DE Mistletoe rMLA protein.
XX
XX Lectin A-chain; mistletoe; rMLA; fusion protein; effector; cytotoxic;
KM intracellular; processing module; protease recognition; targeting module;
KM intracellular; treatment; disorder; cell proliferation; activation;
KM internalisation; allergy; tumour; ricin; translocation.
XX
XX autoimmune disease; allergy; tumour; ricin; translocation.
XX
XX Viscum album.
OS
XX MO9829540-A2.
PN
XX
XX 09-JUL-1998.
PD
XX
XX 02-JAN-1998; 98WO-EP00009.
PE
XX
XX 02-JAN-1997; 97EP-0100012.
PR
XX
XX (BRA1-) BRAIN BIOTECHNOLOGY RES & INFORMATION NE.
PA
XX
XX Eck J, Schmidt A, Zinke H;
PI
XX
XX WPI: 1998-388122/33.
DR
XX
XX N-PSDB; AAV51341.
PT
XX
XX Nucleic acid encoding fusion protein containing mistletoe lectin A
PT chain - useful for treatment of proliferative and autoimmune
PT diseases, allergies and tumours
XX
XX
XX Disclosure; Fig 11a; 115pp; German.
XX
XX
XX This sequence represents a lectin A-chain, rMLA, isolated from mistletoe.
CC This sequence can be used in the construction of a fusion protein which
CC comprises an effector module that is cytotoxic intracellularly, a
CC processing module covalently bonded to the effector module and
CC containing a protease recognition sequence, and a targeting module
CC covalently bonded to the processing module, able to bind specifically to
CC the surface of a cell so as to mediate internalisation of the fusion
CC protein. Such a fusion protein can be used for treating disorders
CC involving proliferation and/or elevated activation of cells, especially
CC autoimmune disease, allergy and tumours. The proteins can be administered
CC e.g. by injection or topically but especially by intravenous injection,
CC at 1 ng to 500 mu g/kg/day, or for ex vivo use at 1 pg to 500 ng/mL.
CC Fusion proteins can develop toxic activity in a wide range of target
CC cells. The processing module prevents extracellular dissociation, and
CC fusion proteins based on mistletoe lectin A-chain are far more active
CC than those based on ricin and do have the associated problems of
CC non-specific toxicity. The protein may be expressed in a non-glycosylated
CC form that does not bind to sugar receptors in the liver, and which has a
CC long half-life in the blood. Where the mistletoe lectin B-chain is used,
CC it actively assists in translocation of the ML A-chain from the
CC endoplasmic reticulum to the cytoplasm.
XX
XX
XX Sequence 252 AA;
SQ
XX
XX Query Match 71.9%; Score 110; DB 19; Length 252;
XX Best Local Similarity 76.9%; Pred. No. 4.2e-09;
XX Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
OY 1 YEREKLRVTHQTGDQYFKFTLLAD 26
DB 2 YERIRLRVTHQTGGEYFRFTLLRD 27

```

```

KM Lectin A-chain; mistletoe; rMLA; fusion protein; effector; cytotoxic;
KM intracellular; processing module; protease recognition; targeting module;
KM intracellular; treatment; disorder; cell proliferation; activation;
KM autoimmune disease; allergy; tumour; ricin; translocation; ss.
XX
XX Viscum album.
XX
XX
XX Key Location/Qualifiers
XX Protein 1..252
XX /note= "partial"
XX
XX MO9829540-A2.
PN
XX
XX 09-JUL-1998.
PD
XX
XX 02-JAN-1998; 98WO-EP00009.
PE
XX
XX 02-JAN-1997; 97EP-0100012.
PR
XX
XX (BRA1-) BRAIN BIOTECHNOLOGY RES & INFORMATION NE.
PA
XX
XX Eck J, Schmidt A, Zinke H;
PI
XX
XX WPI: 1998-388122/33.
DR
XX
XX N-PSDB; AAV51343.
PT
XX
XX Nucleic acid encoding fusion protein containing mistletoe lectin A
PT chain - useful for treatment of proliferative and autoimmune
PT diseases, allergies and tumours
XX
XX
XX Disclosure; Fig 11a'; 115pp; German.
XX
XX
XX This sequence encodes a variant mistletoe lectin A-chain, rMLA. This
CC sequence can be used in the construction of a fusion protein which
CC comprises an effector module that is cytotoxic intracellularly, a
CC processing module covalently bonded to the effector module and
CC containing a protease recognition sequence, and a targeting module
CC covalently bonded to the processing module, able to bind specifically to
CC the surface of a cell so as to mediate internalisation of the fusion
CC protein. Such a fusion protein can be used for treating disorders
CC involving proliferation and/or elevated activation of cells, especially
CC autoimmune disease, allergy and tumours. The proteins can be administered
CC e.g. by injection or topically but especially by intravenous injection,
CC at 1 ng to 500 mu g/kg/day, or for ex vivo use at 1 pg to 500 ng/mL.
CC Fusion proteins can develop toxic activity in a wide range of target
CC cells. The processing module prevents extracellular dissociation, and
CC fusion proteins based on mistletoe lectin A-chain are far more active
CC than those based on ricin and do have the associated problems of
CC non-specific toxicity. The protein may be expressed in a non-glycosylated
CC form that does not bind to sugar receptors in the liver, and which has a
CC long half-life in the blood. Where the mistletoe lectin B-chain is used,
CC it actively assists in translocation of the ML A-chain from the
CC endoplasmic reticulum to the cytoplasm.
XX
XX
XX Sequence 252 AA;
SQ
XX
XX Query Match 71.9%; Score 110; DB 19; Length 252;
XX Best Local Similarity 76.9%; Pred. No. 4.2e-09;
XX Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
OY 1 YEREKLRVTHQTGDQYFKFTLLAD 26
DB 1 YERIRLRVTHQTGGEYFRFTLLRD 26

```

```

DE Prepro mistletoe lectin A chain.
XX
KM Mistletoe; lectin; cytotoxic; A chain; B chain; dimer.
XX
OS Viscum album.
XX
PN EP751221-A1.
XX
PD 02-JAN-1997.
XX
PF 26-JUN-1995; 95EP-0109949.
XX
PR 26-JUN-1995; 95EP-0109949.
XX
PA (MADU ) MADAUUS KOELN AG.
XX
PI Baur A, Eck J, Lentzen H, Zinke H;
XX
DR WPI; 1997-054678/06.
XX
DR N-PSDB; AAT70474.
XX
XX Nucleic acid encoding pre-pro form of mistletoe lectin - for
XX therapeutic or diagnostic use
XX
PS Claim 12; Fig 4A; 30pp; German.
XX
CC Mistletoe lectin is a cytotoxic agent that has been used for tumour
CC therapy. It can be used in immunotoxins and medicaments. Nucleic
CC acid fragments can be used in diagnostic methods. Mistletoe lectin (
CC AAT70473) comprises an A chain (AAT70474) and a B chain (AAT70475).
XX
SQ Sequence 253 AA;

Query Match 71.9%; Score 110; DB 18; Length 253;
Best Local Similarity 76.9%; Pred. No. 4.2e-09;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 YEREKLRVTHQTGDOYFKFTLLAD 26
   |||:|||||:|||||:|||||
Db 2 YERIRLRVTHQTGEYFRFTLLRD 27

RESULT 12
AAW90125
ID AAW90125 standard; Protein; 253 AA.
XX
AC AAW90125;
XX
DT 30-APR-1999 (first entry)
XX
DE Mistletoe ML A-chain protein.
XX
KM ML; mistletoe; lectin; MUA; A-chain; transgenic plant; glycosylation;
KM dimer; immunotoxin; large-scale production; diagnosis; therapeutic;
KM cancer.
XX
XX Viscum album.
OS
XX
XX EP84388-A1.
XX
XX 16-DEC-1998.
XX
XX 26-JUN-1995; 95EP-0109949.
XX
XX 26-JUN-1995; 95EP-0109949.
XX
XX 26-JUN-1995; 98EP-0105660.
XX
XX (MADU ) MADAUUS KOELN AG.
XX
XX Baur A, Eck J, Lentzen H, Zinke H;
XX
XX WPI; 1999-026582/03.
XX
XX N-PSDB; AAT74180.

```

```

XX
XX New transgenic plant expressing mistletoe lectin - useful for
PT producing recombinant lectin in e.g. cancer diagnosis and therapy
XX
XX Disclosure; Fig 4a; 30pp; German.
XX
PS This invention describes a novel transgenic plant transformed with a
XX vector capable of encoding a mistletoe (Viscum album) lectin
XX preprotein or a biologically active fragment. The specification
XX also describes a polypeptide produced by a plant where the polypeptide
XX exhibits at least one enzymatic modification other than the glycosylation
XX that occurs in Viscum album or the polypeptide is a fusion protein, a
XX mistletoe lectin polypeptide dimer and an immunotoxin comprising the
XX polypeptide or the polypeptide dimer. The plants are used for large-scale
XX production of mistletoe lectin for diagnostic or therapeutic purposes
XX (e.g. in cancer therapy). This sequence represents the mistletoe lectin
XX A-chain which is contained in expression vector pT7MUA.
XX
SQ Sequence 253 AA;

Query Match 71.9%; Score 110; DB 20; Length 253;
Best Local Similarity 76.9%; Pred. No. 4.2e-09;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 YEREKLRVTHQTGDOYFKFTLLAD 26
   |||:|||||:|||||:|||||
Db 2 YERIRLRVTHQTGEYFRFTLLRD 27

RESULT 13
AAW25980
ID AAW25980 standard; Protein; 254 AA.
XX
AC AAW25980;
XX
DT 18-OCT-1999 (first entry)
XX
DE Mistletoe lectin A1 protein fragment.
XX
KM Mistletoe; lectin; antitumour; immunostimulant; A-chain; MUA; immunity;
KM ribozyme 28S subunit; non-cytotoxic; T-cell activation; immune response;
KM lymphokine-producing macrophage; uncontrolled cell growth; treatment;
KM cancer; cytotoxicity; antigen; isoform; lectin A1.
XX
XX Viscum album.
OS
XX
XX DE19804210-A1.
XX
XX 12-AUG-1999.
XX
XX 03-FEB-1998; 98DE-1004210.
XX
XX 03-FEB-1998; 98DE-1004210.
XX
XX 03-FEB-1998; 98DE-1004210.
XX
XX (BIOSYN ) BIOSYN ARZNEIMITTEL GMBH.
XX
XX Morris P, Stiefel T, Voelter W, Welters P;
XX
XX WPI; 1999-445335/38.
XX
XX N-PSDB; AAZ09104.
XX
XX Preparation of mistletoe lectins in heterologous systems,
XX particularly for use as anticancer agents and immunostimulants
XX
XX Disclosure; Fig 2B; 78pp; German.
XX
XX This invention describes a novel mistletoe lectin (I) and its fragments
XX which have antitumour and immunostimulatory activity. The A-chain (MUA)
XX of the mistletoe lectin binds to, and inactivates, the 28S subunit of
XX ribosomes. Non-cytotoxic forms of (I) activate T-cell and
XX lymphokine-producing macrophages, so stimulate immunity. (I) and its
XX fragments are used to treat uncontrolled cell growth (particularly
XX cancers) and if they lack cytotoxicity, to increase the strength of the

```

CC Immune response, particularly to a co-administered antigen
CC (tumour-associated, bacterial or viral). The method allows production of
CC mistletoe lectin, and its individual chains. In many different isoforms
CC and on a large scale, at any time of the year. Recombinant products are
CC free from toxins present in natural mistletoe extracts. This sequence
CC represents a fragment of the mistletoe lectin A1 protein.
XX

SO Sequence 254 AA;

Query Match 71.9%; Score 110; DB 20; Length 254;

Best Local Similarity 76.9%; Pred. No. 4.2e-09;

Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 YEREKLRVTHQTGDQYFKFTLLAD 26

DB 1 YERLRLRVTHQTGTGEYFRFTLLRD 26

RESULT 14

AAY25983

ID AAY25983 standard; Protein; 254 AA.

XX

AC AAY25983;

XX

DT 18-OCT-1999 (first entry)

XX

DE Mistletoe lectin A1 (variant) protein fragment.

XX

KW Mistletoe; lectin; antitumour; immunostimulant; A-chain; MIA; immunity;
KW ribozyme 28S subunit; non-cytotoxic; T-cell activation; immune response;
KW lymphokine-producing macrophage; uncontrolled cell growth; treatment;
KW cancer; cytotoxicity; antigen; isoform; lectin A1.

XX

XX

OS Viscum album.

XX

PN DE19804210-A1.

XX

PD 12-AUG-1999.

XX

PF 03-FEB-1998; 98DE-1004210.

XX

PR 03-FEB-1998; 98DE-1004210.

XX

PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.

XX

PI Morris P, Stiefel T, Voelter W, Welters P;

XX

DR WPI; 1999-445335/38.

XX

DR N-PSDB; AA209107.

XX

XX

PT Preparation of mistletoe lectins in heterologous systems,

XX

PT particularly for use as anticancer agents and immunostimulants

XX

PS Disclosure; Fig 5B; 78pp; German.

XX

XX

CC This invention describes a novel mistletoe lectin (I) and its fragments

CC which have antitumour and immunostimulatory activity. The A-chain (MIA)

CC of the mistletoe lectin binds to, and inactivates, the 28S subunit of

CC ribosomes. Non-cytotoxic forms of (I) activate T-cell and

CC lymphokine-producing macrophages, so stimulate immunity. (I) and its

CC fragments are used to treat uncontrolled cell growth (particularly

CC cancers) and if they lack cytotoxicity, to increase the strength of the

CC immune response, particularly to a co-administered antigen

CC (tumour-associated, bacterial or viral). The method allows production of

CC mistletoe lectin, and its individual chains. In many different isoforms

CC and on a large scale, at any time of the year. Recombinant products are

CC free from toxins present in natural mistletoe extracts. This sequence

CC represents a fragment of a mistletoe lectin A1 protein variant.

XX

SO Sequence 254 AA;

Query Match 71.9%; Score 110; DB 20; Length 254;

Best Local Similarity 76.9%; Pred. No. 4.2e-09;

Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 YEREKLRVTHQTGDQYFKFTLLAD 26

DB 1 YERLRLRVTHQTGTGEYFRFTLLRD 26

RESULT 15

AAY25979

ID AAY25979 standard; Protein; 531 AA.

XX

AC AAY25979;

XX

DT 18-OCT-1999 (first entry)

XX

DE Mistletoe lectin I protein fragment.

XX

KW Mistletoe; lectin; antitumour; immunostimulant; A-chain; MIA; immunity;

KW ribozyme 28S subunit; non-cytotoxic; T-cell activation; immune response;

KW lymphokine-producing macrophage; uncontrolled cell growth; treatment;

KW cancer; cytotoxicity; antigen; isoform; lectin I.

XX

OS Viscum album.

XX

PN DE19804210-A1.

XX

PD 12-AUG-1999.

XX

PF 03-FEB-1998; 98DE-1004210.

XX

PR 03-FEB-1998; 98DE-1004210.

XX

PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.

XX

PI Morris P, Stiefel T, Voelter W, Welters P;

XX

DR WPI; 1999-445335/38.

XX

DR N-PSDB; AA209103.

XX

XX

PS Claim 7; Fig 1B; 78pp; German.

XX

CC This invention describes a novel mistletoe lectin (I) and its fragments

CC which have antitumour and immunostimulatory activity. The A-chain (MIA)

CC of the mistletoe lectin binds to, and inactivates, the 28S subunit of

CC ribosomes. Non-cytotoxic forms of (I) activate T-cell and

CC lymphokine-producing macrophages, so stimulate immunity. (I) and its

CC fragments are used to treat uncontrolled cell growth (particularly

CC cancers) and if they lack cytotoxicity, to increase the strength of the

CC immune response, particularly to a co-administered antigen

CC (tumour-associated, bacterial or viral). The method allows production of

CC mistletoe lectin, and its individual chains. In many different isoforms

CC and on a large scale, at any time of the year. Recombinant products are

CC free from toxins present in natural mistletoe extracts. This sequence

CC represents a mistletoe lectin I protein fragment.

XX

SO Sequence 531 AA;

Query Match 71.9%; Score 110; DB 20; Length 531;

Best Local Similarity 76.9%; Pred. No. 1.1e-08;

Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 YEREKLRVTHQTGDQYFKFTLLAD 26

DB 1 YERLRLRVTHQTGTGEYFRFTLLRD 26

Search completed: January 29, 2003, 06:45:56
Job time : 33.5439 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 06:44:49 ; Search time 11.7018 Seconds
(without alignments)
72.918 Million cell updates/sec

Title: US-09-627-165D-19

Perfect score: 153

Sequence: 1 YEREKLRVTHQTGQYFKFYLTDQHS 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	114	74.5	564 4 US-08-776-059-35	Sequence 35, Appl
2	110	71.9	253 4 US-08-776-059-31	Sequence 31, Appl
3	78	51.0	50 4 US-08-776-059-37	Sequence 37, Appl
4	74	48.4	18 4 US-08-776-059-52	Sequence 52, Appl
5	45	29.4	1562 3 US-09-320-878-3	Sequence 3, Appl
6	45	29.4	1562 3 US-09-105-537-35	Sequence 35, Appl
7	45	29.4	11877 4 US-09-105-537-6	Sequence 6, Appl
8	44	28.8	462 3 US-09-238-796-2	Sequence 2, Appl
9	41.5	27.1	127 4 US-08-858-207A-414	Sequence 414, App
10	41	26.8	321 4 US-09-254-465A-2	Sequence 2, Appl
11	41	26.8	545 4 US-08-687-590-26	Sequence 26, Appl
12	41	26.8	584 4 US-08-448-196A-7	Sequence 25, Appl
13	41	26.8	639 4 US-09-422-869-25	Sequence 11, Appl
14	41	26.8	747 4 US-09-362-336A-14	Sequence 4, Appl
15	41	26.8	791 3 US-08-537-361E-4	Sequence 4, Appl
16	41	26.8	791 4 US-08-817-707-4	Sequence 2, Appl
17	41	26.8	792 1 US-08-326-670A-2	Sequence 2, Appl
18	41	26.8	792 3 US-08-537-361E-6	Sequence 6, Appl
19	41	26.8	792 3 US-08-537-361E-5	Sequence 2, Appl
20	41	26.8	792 3 US-08-990-470A-2	Sequence 2, Appl
21	41	26.8	792 4 US-08-817-707-2	Sequence 1, Appl
22	41	26.8	4551 3 US-09-320-878-1	Sequence 31, Appl
23	41	26.8	4613 4 US-09-105-537-31	Sequence 9, Appl
24	40.5	26.5	286 5 PCT-US92-00282-9	Sequence 5, Appl
25	40.5	26.5	531 5 PCT-US92-00282-5	Sequence 2, Appl
26	40.5	26.5	1239 2 US-08-937-931-2	Sequence 2, Appl
27	40.5	26.5	1239 4 US-09-285-502-2	Sequence 2, Appl

28	40.5	26.5	1239 4 US-09-709-126-2	Sequence 2, Appl
29	40.5	26.5	1239 4 US-09-871-385A-2	Sequence 2, Appl
30	40	26.1	219 4 US-08-914-375C-10	Sequence 10, Appl
31	40	26.1	1055 2 US-08-659-251-5	Sequence 5, Appl
32	40	26.1	1055 5 US-09-256-490-5	Sequence 5, Appl
33	40	26.1	1055 5 PCT-US96-11445-5	Sequence 10, Appl
34	40	26.1	3724 2 US-08-804-227C-10	Sequence 4, Appl
35	40	26.1	3724 2 US-08-804-198-4	Sequence 52, Appl
36	40	26.1	4544 1 US-08-469-486-52	Sequence 2, Appl
37	40	26.1	4544 2 US-08-469-486-52	Sequence 447, App
38	39.5	25.8	971 4 US-09-405-728-2	Sequence 2, Appl
39	39	25.5	64 4 US-08-936-165A-447	Sequence 63, Appl
40	39	25.5	111 4 US-09-325-932A-63	Sequence 6, Appl
41	39	25.5	240 2 US-08-380-403A-6	Sequence 6, Appl
42	39	25.5	240 2 US-08-895-628-6	Sequence 6, Appl
43	39	25.5	240 4 US-08-895-810D-6	Sequence 6, Appl
44	39	25.5	263 5 PCT-US94-00844-11	Sequence 11, Appl
45	39	25.5	303 4 US-09-134-001C-4369	Sequence 4369, Ap

ALIGNMENTS

```
RESULT 1
US-08-776-059-35
; Sequence 35, Application US/08776059B
; Patent No. 6271368
; GENERAL INFORMATION:
; APPLICANT: LENTZEN, Hans
; APPLICANT: ECK, Jurgen
; APPLICANT: BAUR, Axel
; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)
; FILE REFERENCE: 674503-2003
; CURRENT APPLICATION NUMBER: US/08/776,059B
; EARLIER FILING DATE: 1996-06-25
; EARLIER APPLICATION NUMBER: 95109949.8
; EARLIER FILING DATE: 1995-06-26
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 35
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Viscum album
US-08-776-059-35

Query Match      74.5%  Score 114; DB 4; Length 564;
Best Local Similarity 80.8%  Pred. No. 9.3e-11;
Matches 21; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 YEREKLRVTHQTGQYFKFYLTD 26
      |||:|||||||:|||||
DB      34 YERLRVTHQTGQYFKFYLTD 59

RESULT 2
US-08-776-059-31
; Sequence 31, Application US/08776059B
; Patent No. 6271368
; GENERAL INFORMATION:
; APPLICANT: LENTZEN, Hans
; APPLICANT: ECK, Jurgen
; APPLICANT: BAUR, Axel
; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)
; FILE REFERENCE: 674503-2003
; CURRENT APPLICATION NUMBER: US/08/776,059B
; EARLIER FILING DATE: 1999-06-19
; EARLIER APPLICATION NUMBER: PCT/EP96/02273
; EARLIER FILING DATE: 1996-06-25
; EARLIER APPLICATION NUMBER: 95109949.8
```

```
; EARLIER FILING DATE: 1995-06-26
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Viscum album
US-08-776-059-31
```

```
Query Match
Best Local Similarity 71.9%; Score 110; DB 4; Length 253;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 YEREKLRVTHQTTGQYFKFTILLAD 26
    ||| :|||||:|||||:|||||
Db 2 YERIRLRVTHQTTGGEYFFRFTILLAD 27
```

```
RESULT 3
US-08-776-059-37
; Sequence 37, Application US/08776059B
; Patent No. 6271368
; GENERAL INFORMATION:
; APPLICANT: LENTZEN, Hans
; APPLICANT: ECK, Jurgen
; APPLICANT: BAUR, Axel
; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)
; FILE REFERENCE: 674503-2003
; CURRENT APPLICATION NUMBER: US/08/776,059B
; EARLIER FILING DATE: 1999-06-19
; EARLIER APPLICATION NUMBER: PCT/EP96/02273
; EARLIER FILING DATE: 1996-06-25
; EARLIER APPLICATION NUMBER: 95109949.8
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Viscum album
US-08-776-059-37
```

```
Query Match
Best Local Similarity 51.0%; Score 78; DB 4; Length 50;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 YEREKLRVTHQTTGQY 17
    ||| :|||||:|||||
Db 34 YERIRLRVTHQTTGDEX 50
```

```
RESULT 4
US-08-776-059-52
; Sequence 52, Application US/08776059B
; Patent No. 6271368
; GENERAL INFORMATION:
; APPLICANT: LENTZEN, Hans
; APPLICANT: ECK, Jurgen
; APPLICANT: BAUR, Axel
; APPLICANT: ZINKE, Holger
; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)
; FILE REFERENCE: 674503-2003
; CURRENT APPLICATION NUMBER: US/08/776,059B
; EARLIER FILING DATE: 1999-06-19
; EARLIER APPLICATION NUMBER: PCT/EP96/02273
; EARLIER FILING DATE: 1996-06-25
; EARLIER APPLICATION NUMBER: 95109949.8
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 52
; LENGTH: 18
```

```
; TYPE: PRT
; ORGANISM: Spodoptera frugiperda
US-08-776-059-52
```

```
Query Match
Best Local Similarity 48.4%; Score 74; DB 4; Length 18;
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 YEREKLRVTHQTTGQY 17
    ||| :|||||:|||||
Db 2 YERIRLRVTHQTTGEY 18
```

```
RESULT 5
US-09-320-878-3
; Sequence 3, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; EARLIER FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-3
```

```
Query Match
Best Local Similarity 29.4%; Score 45; DB 3; Length 1562;
Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 1;
```

```
OY 2 EREKL--RVTHQTTGQYFKFTILLAD 26
    ||| | ||| : ||| : |||
Db 970 DREALARLRALTGSGFGVSVLLDD 996
```

```
RESULT 6
US-09-105-537-35
; Sequence 35, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/105,537A
; EARLIER FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 1562
```



```
; TYPE: PR1
; ORGANISM: Streptomyces venezuelae
US-09-105-537-35

Query Match
Best Local Similarity 44.4%; Score 45; DB 4; Length 1562;
Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 1;

OY 2 EREKL--RVTHQTTGDOYFFITLLAD 26
    ||| ||| ||| : ||| |
Db 970 DREALARLALTGTGDFGVSVLLDD 996

RESULT 7
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438951
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PR1
; ORGANISM: Streptomyces venezuelae
US-09-105-537-6

Query Match
Best Local Similarity 44.4%; Score 45; DB 4; Length 11877;
Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 1;

OY 2 EREKL--RVTHQTTGDOYFFITLLAD 26
    ||| ||| ||| : ||| |
Db 9658 DREALARLALTGTGDFGVSVLLDD 9684

RESULT 8
US-09-238-796-2
; Sequence 2, Application US/09238796
; Patent No. 6074845
; GENERAL INFORMATION:
; APPLICANT: Aiyar, Nambi V.
; APPLICANT: Disa, Jyoti
; TITLE OF INVENTION: BECCR:BOVINE CALCITONIN RECEPTOR-LIKE
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,796
; FILING DATE: 28-JAN-1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-70599
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-238-796-2

Query Match
Best Local Similarity 47.1%; Score 44; DB 3; Length 462;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 5 KLRVTHQTTGDOYFFI 21
    ||| ||| : ||| |
Db 320 KLRVTHQESNLNMYKAV 336

RESULT 9
US-08-858-207A-414
; Sequence 414, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Scodola, Robert
; TITLE OF INVENTION: No. 6348328e1 Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimm1, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 414:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 127 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-414
```

MOLECULE TYPE: protein

```

;      HYPOTHETICAL: NO
;      ANTI-SENSE: NO
;      FRAGMENT TYPE: N-terminal
US-08-448-196A-7

```

Query Match	26.88;	Score 41;	DB 1;	Length 584;
Best Local Similarly	34.68;	Pred. No. 1.3e+02;		
Matches	9;	Conservative	6;	Mismatches 9;
			Indels	2;
			Gaps	1;

```
QY      2 EREKLRVTH--QTGGDQYFKFITLLA 25
      | | : | : | : | : | : |
Db      1 EAHKSEIAHREKFDLGEQHFKGLVLA 26
```

US-09-422-869-25
US-09-422-869-25

RESULT 13

Sequence 25, Application US/09422869

Patent No. 6235481

GENERAL INFORMATION:

APPLICANT: POLONSKY, KENNETH S.

APPLICANT: HORIKAWA, YUKIO

APPLICANT: ODA, NAOKI SA

APPLICANT: COX, NANCY J.

APPLICANT: GREENAN, SEAMUS

APPLICANT: ZHOU, YUN-PING

APPLICANT: OTANI, KENICHI

APPLICANT: HANTS, CRAIG L.

APPLICANT: BELL, GRAEME I.

TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES

FILE REFERENCE: ARCD.307

CURRENT APPLICATION NUMBER: US/09/422, 869

CURRENT FILING DATE: 1999-10-21

EARLIER APPLICATION NUMBER: 60/134, 175

EARLIER FILING DATE: 1999-05-13

NUMBER OF SEQ ID NOS: 30

SOFTWARE: SeqIdN Ver. 2.0

SEQ ID NO 25

LENGTH: 639

TYPE: PRT

ORGANISM: Human

US-09-422-869-25

Query Match	26.8%	Score 41	DB 4	Length 639
Best Local Similarity	47.4%	Pred No. 1	4e+02	
Matches 9	Conservative 3	Mismatches 7	Indels 0	Gaps 0
QY	2	EREKLRVTHQTGGDQTEKF	20	
DB	308	EREKMGVTVQDDGDEFMMTF	326	

```

? RESULT 14
? US-09-362-336A-14
? Sequence 14, Application US/09362336A
? Patent No. 6440696
? GENERAL INFORMATION:
? APPLICANT: Bao, Yimla
? TITLE OF INVENTION: No. 6440696el E6 Targeted Protein (ESTP1)
? FILE REFERENCE: 18475-017, ESTP1
? CURRENT APPLICATION NUMBER: US/09/362.336A
? CURRENT FILING DATE: 1999-07-28
? NUMBER OF SEQ ID NOS: 17
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 14
? LENGTH: 747
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: residues 104-850 of Genbank Accession No. 6440696
US-09-362-336A-14

```

```

Best Local Similarity 34.8%; Pred. No. 1.7e+02;
Matches 8; Conservative 5; Mismatches 10; Indels 0; Gaps 0.
QY 4 EKLRVTHQFTGDYFKEKITLAD 26
| : : | : : | | | |
Db 244 EEEMYNNOEGRAFMQFTLLGD 266

```

```
QY      4  EKL RVTHQTGDQYFKFITLLAD 26
          | :  : |  :  : | | | | |
Db      244 EEE MYNNQ EAGPAFMQFLTLLGD 266
```

```

RESULT 15
US-08-537-361E-4
Sequence 4, Application US/08537361E
Patent No. 6121037
GENERAL INFORMATION:
APPLICANT: Stojiljkovic, Igor
APPLICANT: So, Magdalene
APPLICANT: Hwa, Vivian
APPLICANT: Heffron, Fred
APPLICANT: Nassif, Xavier
TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
TITLE OF INVENTION: Genes and Uses
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,361E
FILING DATE: 02-OCT-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 6121037nan, Kevin E
REGISTRATION NUMBER: 35,103
REFERENCE/DOCKET NUMBER: 94,784-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 791 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-537-361E-4

```

Query Match Similarity	26.8%	Score 41:	DB 3:	Length 791:
Best Local Similarity	32.1%	Pred. No. 1.8e+02:		
Matches	9:	Conservative	6:	Mismatches 11:
			Indels	0:
			Gaps	0:
QY	1	YEREKLRVTHQTGGQYEFKFTLLADQH	28	
			
Db	354	YHRKEVGELYNRSMDTEFRKRIILRMDSH	381	

Search completed: January 29, 2003, 06:47:47
Job time : 15.7018 secs

Search completed: January 29, 2003, 06:47:47
Job time : 15.7018 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 06:46:04 ; Search time 8.14035 Seconds
(without alignments)
71.886 Million cell updates/sec

Title: US-09-627-165D-19
Perfect score: 153
Sequence: 1 YERKRLVHTQTGQDYKFKFTLLADQHS 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications-AA:*
1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCF_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/PCFUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/PCFUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	71.9	252	10 US-09-347-064-2	Sequence 2, Appli
2	110	71.9	252	10 US-09-347-064-8	Sequence 8, Appli
3	45	29.4	1562	9 US-09-860-846-35	Sequence 35, Appli
4	45	29.4	1562	10 US-09-861-289-35	Sequence 35, Appli
5	45	29.4	11877	9 US-09-860-846-6	Sequence 6, Appli
6	45	29.4	11877	10 US-09-861-289-6	Sequence 6, Appli
7	44.5	29.1	669	9 US-09-983-204-15	Sequence 15, Appli
8	44.5	29.1	669	10 US-10-133-157-4	Sequence 4, Appli
9	44	28.8	72	10 US-09-864-761-34449	Sequence 34449, A
10	44	28.8	125	10 US-09-864-761-47239	Sequence 47239, A
11	44	28.8	461	9 US-10-051-186-2	Sequence 2, Appli
12	44	28.8	464	9 US-10-051-186-9	Sequence 9, Appli
13	43	28.1	195	9 US-10-074-956-23	Sequence 23, Appli
14	43	28.1	241	9 US-10-074-956-29	Sequence 29, Appli
15	42.5	27.8	253	9 US-09-834-406-7	Sequence 7, Appli
16	42	27.5	251	9 US-09-895-912A-90	Sequence 90, Appli
17	41	26.8	306	10 US-09-881-752A-242	Sequence 242, App
18	41	26.8	321	9 US-09-978-295A-52	Sequence 52, Appli
19	41	26.8	321	9 US-09-978-697-52	Sequence 52, Appli

20	41	26.8	321	9 US-09-978-192A-52	Sequence 52, Appli
21	41	26.8	321	9 US-09-953-499-2	Sequence 2, Appli
22	41	26.8	321	9 US-09-999-832A-52	Sequence 52, Appli
23	41	26.8	321	9 US-09-978-189-52	Sequence 52, Appli
24	41	26.8	396	9 US-09-738-626-6216	Sequence 6216, Ap
25	41	26.8	429	10 US-09-866-582-12	Sequence 12, Appli
26	41	26.8	559	9 US-09-854-133-396	Sequence 396, App
27	41	26.8	559	10 US-09-738-873-396	Sequence 396, App
28	41	26.8	639	10 US-09-768-877-25	Sequence 25, Appli
29	41	26.8	613	9 US-09-860-846-31	Sequence 31, Appli
30	41	26.8	4613	10 US-09-861-289-31	Sequence 31, Appli
31	40.5	26.5	275	10 US-09-815-242-12069	Sequence 12069, A
32	40.5	26.5	287	10 US-09-305-856B-10	Sequence 10, Appli
33	40.5	26.5	1239	10 US-09-871-388-2	Sequence 2, Appli
34	40	26.1	110	10 US-09-864-761-44148	Sequence 44148, A
35	40	26.1	166	10 US-09-764-853-717	Sequence 717, App
36	40	26.1	178	10 US-09-764-853-469	Sequence 469, App
37	40	26.1	258	9 US-09-813-453A-6	Sequence 6, Appli
38	40	26.1	307	10 US-09-816-714-2	Sequence 2, Appli
39	40	26.1	428	12 US-10-042-417-6	Sequence 6, Appli
40	40	26.1	429	10 US-09-866-582-30	Sequence 30, Appli
41	40	26.1	766	10 US-09-925-301-1382	Sequence 1382, Ap
42	40	26.1	888	10 US-09-810-786-5	Sequence 5, Appli
43	40	26.1	897	10 US-09-810-786-4	Sequence 4, Appli
44	40	26.1	923	10 US-09-825-147-2	Sequence 2, Appli
45	40	26.1	932	10 US-09-813-148-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-347-064-2
; Sequence 2, Application US/09347064A
; Patent No. US20020045208A1
; GENERAL INFORMATION:
; APPLICANT: Eck, Jürgen
; APPLICANT: Schmidt, Arno
; APPLICANT: Zinke, Holger
; TITLE OF INVENTION: Recombinant Fusion Proteins Based on
; TITLE OF INVENTION: Ribosome-Inactivating Proteins of the mistletoe Viscum
; FILE REFERENCE: 09282-5
; CURRENT APPLICATION NUMBER: US/09/347,064A
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: PCT/EP98/00009
; EARLIER FILING DATE: 1998-01-02
; EARLIER APPLICATION NUMBER: EP 97 10 0012.0
; EARLIER FILING DATE: 1997-01-02
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Viscum album
US-09-347-064-2
Query Match 71.9%; Score 110; DB 10; Length 252;
Best local Similarity 76.9%; Pred. No. 3.4e-10;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
OY 1 YERKRLVHTQTGQDYKFKFTLLAD 26
|||:|||||||:|||||
Db 2 YERIRLVRHTQTGEYKFTLLRD 27
RESULT 2
US-09-347-064-8
; Sequence 8, Application US/09347064A
; Patent No. US20020045208A1
; GENERAL INFORMATION:
; APPLICANT: Eck, Jürgen
; APPLICANT: Schmidt, Arno

```
; APPLICANT: Zinke, Holger
; TITLE OF INVENTION: Recombinant Fusion Proteins Based on
; TITLE OF INVENTION: Ribosome-Inactivating Proteins of the mistletoe Viscum
; FILE REFERENCE: album
; CURRENT APPLICATION NUMBER: US/09/347,064A
; EARLIER FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: PCT/EP98/00009
; EARLIER FILING DATE: 1998-01-02
; EARLIER APPLICATION NUMBER: EP 97 10 0012.0
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Viscum album
US-09-347-064-8
```

```
Query Match          71.9%; Score 110; DB 10; Length 252;
Best Local Similarity 76.9%; Pred. No. 3,4e-10;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 YEREKLRVTHQTGGDYFKFTLLAD 26
Db 1 YERIRLRVTHQTGGDYFKFTLLAD 26
```

```
RESULT 3
; Sequence 35, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-35
```

```
Query Match          29.4%; Score 45; DB 9; Length 1562;
Best Local Similarity 44.4%; Pred. No. 73;
Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 1;
```

```
OY 2 EREKL--RVTHQTGGDYFKFTLLAD 26
Db 970 DREALARLRLTGTGGFTGVSLDD 996
```

```
RESULT 4
; Sequence 35, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
```

```
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-35
```

```
Query Match          29.4%; Score 45; DB 10; Length 1562;
Best Local Similarity 44.4%; Pred. No. 73;
Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 1;
```

```
OY 2 EREKL--RVTHQTGGDYFKFTLLAD 26
Db 970 DREALARLRLTGTGGFTGVSLDD 996
```

```
RESULT 5
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-6
```

```
Query Match          29.4%; Score 45; DB 9; Length 11877;
Best Local Similarity 44.4%; Pred. No. 8.3e+02;
Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 1;
```

```
OY 2 EREKL--RVTHQTGGDYFKFTLLAD 26
Db 9658 DREALARLRLTGTGGFTGVSLDD 9684
```

```
RESULT 6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-6
```


OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.8
OTHER INFORMATION: EST HUMAN HIT: AA298652.1, EVALUATE 2.00e-34
OTHER INFORMATION: SWISSPROT HIT: Q16602, EVALUATE 8.00e-36
US-09-864-761-34449

Query Match

Best Local Similarity 28.8%; Score 44; DB 10; Length 72;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 5 KLRVTHQTTDQYFKF1 21
Db 16 KLRVTHQESNLYMKAV 32

RESULT 10

US-09-864-761-47239
Sequence 47239, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm103-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263, 6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 47239
LENGTH: 125
TYPE: prt
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006115.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.58

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.49
OTHER INFORMATION: EST HUMAN HIT: AM873518.1, EVALUATE 3.00e-19
OTHER INFORMATION: SWISSPROT HIT: Q90L58, EVALUATE 9.00e-26
US-09-864-761-47239

Query Match

Best Local Similarity 28.8%; Score 44; DB 10; Length 125;
Matches 9; Conservative 7; Mismatches 4; Indels 2; Gaps 1;

QY 2 EREKRVTHQTTDQYFKF1 23
Db 32 EREKRVTHQESNLYMKAV 51

RESULT 11

US-10-051-186-2
Sequence 2, Application US/10051186
Patent No. US20020164707A1
GENERAL INFORMATION:
APPLICANT: Adamou, J., et al.
TITLE OF INVENTION: Calcitonin Gene Related Peptide Receptor
FILE REFERENCE: PFI29C1
CURRENT APPLICATION NUMBER: US/10/051,186
PRIOR FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: 09/455,442
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 08/461,250
PRIOR FILING DATE: 1995-06-05
PRIOR APPLICATION NUMBER: PCT/US95/01587
PRIOR FILING DATE: 1995-02-03
PRIOR APPLICATION NUMBER: PCT/US94/09235
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 461
TYPE: prt
ORGANISM: Homo sapiens
US-10-051-186-2

Query Match

Best Local Similarity 28.8%; Score 44; DB 9; Length 461;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 5 KLRVTHQTTDQYFKF1 21
Db 319 KLRVTHQESNLYMKAV 335

RESULT 12

US-10-051-186-9
Sequence 9, Application US/10051186
Patent No. US20020164707A1
GENERAL INFORMATION:
APPLICANT: Adamou, J., et al.
TITLE OF INVENTION: Calcitonin Gene Related Peptide Receptor
FILE REFERENCE: PFI29C1
CURRENT APPLICATION NUMBER: US/10/051,186
PRIOR FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: 09/455,442
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 08/461,250
PRIOR FILING DATE: 1995-06-05
PRIOR APPLICATION NUMBER: PCT/US95/01587
PRIOR FILING DATE: 1995-02-03
PRIOR APPLICATION NUMBER: PCT/US94/09235
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 9
LENGTH: 464
TYPE: prt
ORGANISM: Rat

1

Gencore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 06:44:49 ; Search time 14.2456 Seconds
(without alignments)
195.702 Million cell updates/sec

Title: US-09-627-165d-19
Perfect score: 153
Sequence: 1 YERKLRVTHQTGDDYFKRFTLLADQHS 29

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	71.9	254	2 PD0018	mistletoe lectin I
2	51.5	33.7	454	2 S77005	sensory transducti
3	51	33.3	464	2 T20238	hypothetical prote
4	50	32.7	161	2 AD2540	hypothetical prote
5	49.5	32.4	1147	1 MMAX1B	myosin heavy chain
6	49	32.0	261	2 S12706	type II site-speci
7	48.5	31.7	510	2 T16146	hypothetical prote
8	48.5	31.7	1168	1 MMAX1C	myosin heavy chain
9	48	31.4	50	2 H97531	hypothetical prote
10	48	31.4	190	2 F69336	replication contro
11	48	30.4	318	2 AC2442	hypothetical prote
12	46.5	30.4	841	2 B71212	hypothetical prote
13	46.5	30.4	1075	2 T38253	hypothetical prote
14	46	30.1	381	2 D50884	probable transport
15	46	30.1	381	2 C85734	probable transport
16	46	30.1	700	2 E69146	sensory transducti
17	46	30.1	1100	2 T21544	hypothetical prote
18	45	29.4	88	2 G82781	hypothetical prote
19	45	29.4	140	2 C96957	hypothetical prote
20	45	29.4	174	2 T17753	hypothetical prote
21	45	29.4	373	2 T23000	hypothetical prote
22	45	29.4	383	2 T36080	hypothetical prote
23	45	29.4	488	2 AD0735	probable NADH dehy
24	45	29.4	492	1 S28003	conserved hypotet
25	45	29.4	620	2 T23522	trypanothione-disu
26	45	29.4	701	2 E90202	hypothetical prote
27	45	29.4	1562	2 T17411	ATP synthase subun
28	45	29.4	2358	2 T39569	polyketide synthas
29	45	29.4	2371	2 T43432	probable alpha-glu
					alpha-glucan synth

30	45	29.4	5035	1 T46646	ryanodine receptor
31	44.5	29.1	374	2 D72115	conserved hypotet
32	44.5	29.1	374	2 G86508	hypothetical prote
33	44.5	29.1	605	1 ABPGS	serum albumin prec
34	44.5	29.1	666	1 S64203	ATP-dependent nucl
35	44.5	29.1	669	2 A49585	Na+ channel protel
36	44.5	29.1	1104	2 S36773	GTPase-activating
37	44.5	29.1	1113	1 A47106	myosin heavy chain
38	44.5	29.1	2357	2 A59249	class VII unconven
39	44.5	28.8	119	2 C71556	holo-lacyl-carrier
40	44	28.8	126	2 T12888	hypothetical prote
41	44	28.8	127	2 A82909	transcription term
42	44	28.8	131	2 C71054	hypothetical prote
43	44	28.8	148	2 B72782	hypothetical prote
44	44	28.8	150	2 T44952	flagella-related p
45	44	28.8	172	2 JEO130	acylalone dehydrat

ALIGNMENTS

RESULT 1
PD0018
mistletoe lectin I A chain - Viscum album (fragment)
C:Species: Viscum album
C:Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 26-Aug-1999
C:Accession: PD0018
R:Eschenburg, S.; Krausepenhaar, R.; Mikhailov, A.; Stoeva, S.; Betzel, C.; Voelter, W
Biochem. Biophys. Res. Commun. 247, 367-372, 1998
A:Title: Primary structure and molecular modelling of mistletoe lectin I from Viscum a
A:Reference number: PD0018; M0ID:98308123; PMID:9642133
A:Accession: PD0018
A:Molecule type: protein
A:Residues: 1-254 <ESC>
A:Superfamily: ricin; rRNA N-glycosidase homology
F:7-246/Domain: rRNA N-glycosidase homology <RNG>

Query Match 71.9%; Score 110; DB 2; Length 254;
Best Local Similarity 76.9%; Pred. No. 3.2e-09;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKLRVTHQTGDDYFKRFTLLAD 26
Db 1 YERKLRVTHQTGDEYFRFTLLRD 26

RESULT 2

S77005
sensory transduction histidine kinase s110798 - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein s110798
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 02-Sep-2000

C:Accession: S77005
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocys
S.
A:Reference number: S74322; M0ID:97061201; PMID:8905231
A:Accession: S77005
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-454 <KAN>
A:Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PID:BAA10697.1; PID:g100
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: sensory transduction histidine kinase s110798; sensor histidine kinase
F:203-445/Domain: sensor histidine kinase homology <SHK>

Query Match 33.7%; Score 51.5; DB 2; Length 454;
Best Local Similarity 37.9%; Pred. No. 6.3;
Matches 11; Conservative 6; Mismatches 9; Indels 3; Gaps 1;

```
QY      1 YEREKLRVTHQTGDOYFKFITLLADQHS 29
          : : | : | | | : | : | : | |
Db     134 FNQETWQTIHPPTGDRYRQFTTIL--HS 159
```

RESULT 3
T20238
hypothetical protein C54g4.9 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:accession: T20238

Query Match	33.38	Score 51	DB 2	Length 464
Best Local Similarity	42.98	Pred. No. 7.7		
Matches	9	Conservative	3	Mismatches 9
				Indels 0
				Gaps 0
QY	2	EREKRVTHQTGDTQYFFKFT	22	
Db	427	EEVYMRLEHONIGASYFSILT	447	

RESULT 4
 AD2540
 hypothetical protein all7607 [imported] - Nostoc sp. (strain PCC 7120) plasmid pcc7120b
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AD2540
 R:kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuguhi,
 Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 6, 203-213, 2001
 A:title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD2540
 A:status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-161 <KUP>
 A:Cross-references: GB:AF003602; PIDN:BA877250.1; PID:g17134692; GSPDB:GN00181
 C:Experimental source: strain PCC 7120
 C:Genetics:
 A:gene: all7607
 :genome: plasmid

Query Match	32.7%	Score 50;	DB 2;	Length 161;
Best Local Similarity	32.1%	Pred. No. 3.3;		
Matches	9;	Conservative	8;	Mismatches 11;
				Indels 0;
				Gaps 0;
QY	2	EREKRYVHOTGDOYEFKEFTLLADQHS	29	
		:::		
		:::		
Db	122	EMNNVKRTAATVGTGDKRYFKRISTLTQMA	149	

RESULT 5
MMXIB
myosin heavy chain IB - Acanthamoeba castellanii
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Acanthamoeba castellanii
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990
#text_change 19-Apr-2002
#Accession: J00095; B34448

R:Jung, G.; Schmidt, C.J.; Hammer III, J.A.
Gene 82, 269-280, 1989
A:Title: Myosin I heavy-chain genes of *Acanthamoeba castellanii*: cloning of a second
A:Reference number: J00095; M0ID:90060816; PMID:2511079
A:Accession: J00095

```

r333/alignment site: phosphate (Ser) (covalent) #status predicted
Query Match          32.4%: Score 49.5; DB 1; Length 1147;
Best Local Similarity 53.8%: Pred. NO. 36;
Matches 14; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY      1 YEKRLAVTHQTMGDQYK-FITLLA 25
      ||| ||| ||| ||| |||
Db      179 YLEKSEVVFQTLGERSFHFIFYOLLA 204

```

```

RESULT 6
SI2706      type II site-specific deoxyribonuclease (EC 3.1.21.4) Sau96I - Staphylococcus aureus
N:Alternate names: restriction endonuclease Sau96I
C:Species: Staphylococcus aureus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 24-Nov-1999
C:Accession: SI2706
R:Sezlik, L.; Venetianer, P.; Kiss, A.
Nucleic Acids Res. 18, 4659-4664, 1990
A:Title: Cloning and nucleotide sequence of the genes coding for the Sau96I restriction
A:Reference number: SI2705; MUID:90370441; PMID:2204026
A:Accession: SI2706
A:Molecule type: DNA
A:Residues: 1-261 <SZ>
A:Cross-references: EMBL:X53096; NID:g46616; PUDN:CAA37259.1; PID:g46618
C:Superfamily: Staphylococcus aureus type II site-specific deoxyribonuclease Sau96I
Keywords: hydrolase

```

Query Match	32.0%	Score 49	DB 2	Length 261
Best Local Similarity	41.7%	Pred. No. 8.1		
Matches	10	Conservative	5	Mismatches 5; Indels 4; Gaps 1
Oy	3	REKLRVTHQ---TTGDQYKFFIT	22	
		:		
		: : :		
Db	172	REKRFSSHPRVRLASGDQFYKIVT	195	

RESULT 7
T16146
hypothetical protein F25B5.6 - *Caenorhabditis elegans*

A:Reference number: AB1807, NCBI:AB1807, PMID:11759840
A:Accession: NC2442
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1318 <KOR>
A:Cross-references: GB:BA000019, PIDN:BA076790.1, PID:g17134229, GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all5091

```

Query Match      31.4%; Score 48; DB 2; Length 318;
Best Local Similarity 45.0%; Pred NO. 14;
Matches      9; Conservative 3; Mismatches 8; Indels 0; Gaps 0
QY      1 YERKLRVYHQTGGDYERF 20
      :||||:|:|:|
Db      94 FERREKALTSQVYGPVILNF 113

```

RESULT 12
B71212
hypothetical protein PH1962 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence,revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: B71212
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekiriri,
M. N.; Ohtuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: B71212
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1841 <RAW>
A:Cross-references: GB:AP000007; NID:G3236134; PIDN:BA31089.1; PIR:G3258406
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1962

```

Query Match 30.4%; Score 46.5; DB 2; Length 841;
Best Local Similarity 32.4%; Pred No. 74;
Matches 12; Conservative 4; Mismatches 12; Indels 9; Gaps 1.

QY      2 EREKLVHTQITGSDQYFK-----TTLADQHS 29
      | | | | | | | | | | | | | | | |
Db      235 ESKKRLANQIQGKQLFSESTEDIQIDILTPRA 271

```

RESULT 13
 T38253
 hypothetical protein SPAC23C11.15 - fission yeast (*Schizosaccharomyces pombe*)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_cvsion 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T38253
 R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, August 1995
 A:Reference number: Z21781
 A:Accession: T38253
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1075 <BRO>
 A:Cross-references: EMBL:Z98559; PIDD:CA81171.1; GSPDB:GN00066; SPDB:SPAC23C11.15
 A:Experimental source: strain 972h-; cosmid c23C11
 C:Genetics:
 A:Gene: SPDB:SPAC23C11.15
 A:Map position: 1
 A:Introns: 11/1; 71/2

Query Match	30.48;	Score 46.5;	DB 2;	Length 1075;
Best Local Similarity	32.48;	Pred. No. 97;		
Matches 11;	Conservative 6;	Mismatches 12;	Indels 5;	Gaps 1;

```
QY 1 YEREKL-----RVTHQTGDQYKFIITLLADQHS 29
      | :|| : | :||:| | :||:
Db 845 YTIDKLWMSAKQVHHIVSDKRYKFEVTSLEQNS 87
```

RESULT 14
D90884
probable transport protein ECs2044 [Imported] - Escherichia coli (strain O157:H7, sub
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90884
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasevarea, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90884
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-381 <HAY>
A:Cross-references: GB:BA000007; PIDN:BAB35467.1; PID:g13361510; GSPDB:GN00154
A:Experimental source: Strain O157:H7, subStrain R1MD 0509952
C:Genetics:
A:Gene: ECs2044

	Query Match	30.1%	Score 46;	DB 2;	Length 381,
	Best Local Similarity	45.0%;	Pred. No. 36;		
	Matches	9; Conservative	4; Mismatches	7; Indels	0; Gaps
Oy	10 HQTGGDYKFKETLLADQHS	29			
	: :: :	: :: :			
Db	211 YLTLEROYQAVLKVRDQHS	230			

RESULT 15
C85734
Probable transport protein 22279 [imported] - *Escherichia coli* (strain O157:H7, subsp. C); Species: *Escherichia coli*
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: C85734
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca Nturu 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85734
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-381 <SMO>
A;Cross-references: GB:AE005174; NID:g12515262; PIDN:AAG56335.1; GSPDB:GN00145; UWGPD:A; Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: 22279

		30.1%;	Score 46;	DB 2;	Length 381;
	Query March				
	Best Local Similarity	45.0%;	Pred. No. 36;		
	Matches	9;	Conservative	4;	Mismatches
				7;	Indels
				0;	Gaps
QY	10 HQTGDDYFFKFTLLADDSH 29				
	: : : : : : :				
Db	211 YQLTEGYCAVINKVLRDDSH 230				

Search completed: January 29, 2003, 06:47:16
Job time : 17.2456 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 06:44:49 ; Search time 18.8246 Seconds
(without alignments)
63.896 Million cell updates/sec

Title: US-09-627-165D-19

Perfect score: 153
Sequence: 1 YERKRLVTHQTTGDIYFKFTTLADQHS 29

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	71.9	254	MLA_VISAL	P81446 viscum albu
2	49.5	32.4	1147	MYSB_ACACA	P19706 acanthamob
3	49	32.0	261	T2S9_STRAU	P23736 staphylococ
4	48.5	31.7	510	FOIC_CAEEL	Q09509 caenorhabdi
5	48.5	31.7	1168	MYSC_ACACA	P10569 acanthamob
6	46.5	30.4	1075	PS72_SCHPO	013919 schizosacch
7	45	29.4	181	Y4WG_RHISN	P55685 rhizobium s
8	45	29.4	492	TYTR_TRYBB	P39051 trypanosoma
9	45	29.4	701	VATI_SULSO	Q9UWV3 sulfolobus
10	45	29.4	2358	MORD_SCHPO	Q97719 schizosacch
11	45	29.4	5035	RTRI_PLG	P16560 sus scrofa
12	44.5	29.1	455	PEX3_PICPA	Q92262 pichia past
13	44.5	29.1	605	ALBU_PLG	P08835 sus scrofa
14	44.5	29.1	666	Y032_MYCGE	P47278 mycoplasma
15	44.5	29.1	669	SCAA_HUMAN	P37088 homo sapien
16	44.5	29.1	1104	BUD2_YEAST	P33313 homo sapien
17	44.5	29.1	1113	MYSD_DICDI	P34109 dictyosteli
18	44	28.8	119	ACPS_CHLTR	084102 chlamydia t
19	44	28.8	150	FLAG_METVO	006640 methanococ
20	44	28.8	172	SCYD_MAGGR	P56221 magnaporthe
21	44	28.8	266	PSB4_YEAST	P30657 saccharomyc
22	44	28.8	461	CGRR_HUMAN	Q16602 homo sapien
23	44	28.8	464	CGRR_RAT	Q63118 rattus norv
24	44	28.8	994	MYSA_DICDI	P22467 dictyosteli
25	44	28.8	1845	YPM8_CAEEL	P34529 caenorhabdi
26	43.5	28.4	503	YPTJ_CAEEL	Q23469 caenorhabdi
27	43.5	28.4	1028	MYIC_HUMAN	Q00159 homo sapien
28	43.5	28.4	1028	MYIC_MOUSE	Q00159 mus musculu
29	43	28.1	292	Y983_AOUAE	Q67109 aquifex aeo
30	43	28.1	608	ALBU_MOUSE	P07724 mus musculu
31	43	28.1	985	AGLU_ASPOR	Q12558 aspergillus
32	43	28.1	3023	POLG_TYVAV	P09814 t genome po
33	42.5	27.8	366	DCX_MOUSE	Q88809 mus musculu

34	42.5	27.8	371	MXIG_SHIFL	Q57332 shigella fl
35	42.5	27.8	402	DCX_HUMAN	Q43602 homo sapien
36	42.5	27.8	740	DCX_HUMAN	Q15075 homo sapien
37	42.5	27.8	756	DCX_MOUSE	Q91188 mus musculu
38	42.5	27.8	863	YEBU_ECOLI	P33924 escherichia
39	42	27.5	238	YFBN_ECOLI	P76484 escherichia
40	42	27.5	250	KDSB_YERPE	Q82944 yersinia pe
41	42	27.5	393	PDNS_YEAST	Q04062 saccharomyc
42	42	27.5	481	TYTR_CRIFA	P39040 crithidia f
43	42	27.5	548	FOLE_YEAST	Q08645 saccharomyc
44	42	27.5	581	YD15_SCHPO	Q10238 schizosacch
45	42	27.5	649	FLR3_HUMAN	Q9NZU0 homo sapien

ALIGNMENTS

RESULT 1	MLA_VISAL	STANDARD:	PRT:	254 AA.
AC	P81446:1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Beta-galactoside specific lectin I A chain (MLA) (ML-I A) (rRNA N-glycosidase) (EC 3.2.2.22).			
DE	Viscum album (European mistletoe).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Santalales; Viscaceae; Viscum.			
OC	Santalales; Viscaceae; Viscum.			
OX	NCBI_TaxID=3972;			
RN	[1]			
RP	SEQUENCE.			
RC	STRAIN-Subsp. album.			
RC	MEDLINE=97134581; PubMed=6980141.			
RA	Huguet Soler M., Stoeva S., Schwaborn C., Wilhelm S., Stiefel T., Voelter W.;			
RT	"Complete amino acid sequence of the A chain of mistletoe lectin I.";			
RL	FEBS Lett. 399:153-157(1996).			
CC	- FUNCTION: THE A CHAIN IS RESPONSIBLE FOR INHIBITING PROTEIN SYNTHESIS THROUGH THE CATALYTIC INACTIVATION OF 60S RIBOSOMAL SUBUNITS BY REMOVING ADENINE FROM POSITION 4,324 OF 28 S RNA. THE B CHAIN BINDS TO CELL RECEPTORS AND PROBABLY FACILITATES THE ENTRY INTO THE CELL OF THE A CHAIN; B CHAINS ARE ALSO RESPONSIBLE FOR CELL AGGLUTINATION (LECTIN ACTIVITY).			
CC	- CATALYTIC ACTIVITY: Endohydrolysis of the N-glycosidic bond at one specific adenosine on the 28S rRNA.			
CC	- SUBUNIT: DISULFIDE-LINKED DIMER OF A AND B CHAINS.			
CC	- MISCELLANEOUS: TWO ISOFORMS OF MLA EXIST: GLYCOSYLATED FORM MLA AND NON-GLYCOSYLATED FORM MLA'.			
CC	- SIMILARITY: BELONGS TO THE RIBOSOME-INACTIVATING PROTEIN FAMILY.			
CC	- TYPE 2 RIP SUBFAMILY.			
DR	HSP: P1140; IABR.			
DR	InterPro: IPR001574; RIP.			
DR	Pfam: PF00161; RIP; 1.			
DR	PRINTS: PR00396; SHIGARICIN.			
DR	PROSITE: PS00275; SHIGA_RICIN; Repeat: Glycoprotein; Lectin.			
KW	Plant defense; Hydrolase; Toxin; Repeat: Glycoprotein; Lectin.			
FT	ACET SITE	165	165	
FT	CAROHND	112	112	
FT	VARIANT	15	15	N-LINKED (GLCNAC. . .).
FT	VARIANT	66	66	E-> D (IN MLA').
FT	VARIANT	112	112	V-> I (IN MLA').
FT	VARIANT	116	116	N-> T (IN MLA').
FT	VARIANT	133	133	P-> T (IN MLA').
FT	VARIANT	140	140	DO-> EE (IN MLA').
FT	VARIANT	144	144	T-> S (IN MLA').
FT	VARIANT	151	151	F-> Y (IN MLA').
FT	VARIANT	179	179	T-> A (IN MLA').
FT	VARIANT	184	184	Y-> D (IN MLA').
FT	VARIANT	190	190	A-> E (IN MLA').
FT	VARIANT	218	218	V-> M (IN MLA').
FT	VARIANT	223	223	I-> F (IN MLA').
FT	VARIANT	224	224	PP-> ST (IN MLA').

FT	VARIANT	231	231	T -> S (IN M.L.A.).
FT	VARIANT	235	235	D -> S (IN M.L.A.).
SQ	SEQUENCE	254 AA:	28478 MW:	53BAF98D30FEF67 CRC64;
<p>Query Match</p> <p>Best Local Similarity 76.9%; Pared No. 8.1e-10;</p> <p>Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0</p>				
QY	1 YERKLRVTHQTTGDDYKFFITLLAD 26			
DB	1 YERLRLRVTHQTTGEEYFFITLLRD 26			
<p>RESULT 2</p> <p>MYSB_ACACA STANDARD; PRT; 1147 AA.</p>				
ID	MYSB_ACACA			
AC	P19706;			
DT	01-FEB-1991 (Rel. 17, Created)			
DR	01-NOV-1995 (Rel. 32, Last sequence update)			
DR	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Myosin heavy chain IB (Myosin heavy chain II).			
GN	MIB OR MIT.			
OS	Acanthamoeba castellanii (Acanthamoeba).			
OC	Eukaryota; Acanthamoebidae; Acanthamoeba.			
OX	NCBI_TaxID=5755;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90060816; PubMed=2511079;			
RA	Jung G., Schmidt C.J., Hammer J.A. III;			
RT	"Myosin I heavy-chain genes of Acanthamoeba castellanii: cloning of a			
RL	second gene and evidence for the existence of a third isoform";			
Gene	82:269-280(1989).			
CC	-1 FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE			
CC	ACTIVITY THAT IS ACTIVATED BY F-ACTIN.			
CC	-1 SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY			
CC	AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.			
CC	-1 DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;			
CC	MYOSINS I COULD THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND			
CC	VIC VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER			
CC	WITH THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE			
CC	MOLECULES OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.			
CC	-1 MISCELLANEOUS: THIS ORGANISM EXPRESSES AT LEAST THREE ISOFORMS OF			
CC	MYOSIN I HEAVY-CHAIN, ENCODED BY GENES MIA, MIB, AND MIC.			
CC	-1 SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.			
CC	-1 SIMILARITY: CONTAINS 1 SH3 DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M30780; AAA27708.1; -			
DR	PIR; J00095; MMAXIB.			
DR	HSSP; P08799; 1MND.			
DR	InterPro; IPR000048; IQ_region.			
DR	InterPro; IPR001452; SH3.			
DR	InterPro; IPR001609; myosin_head.			
DR	Pfam; PF00018; SH3; 1.			
DR	Pfam; PF00063; myosin_head; 1.			
DR	Pfam; PF00612; IQ; 1.			
DR	PRINTS; PR00193; MYOSINHEAVY.			
DR	PRINTS; PR00452; SH3DOMAIN.			
DR	PRODom; PD000066; SH3; 1.			
DR	PRODom; PD000355; myosin_head; 1.			
DR	SMART; SM00242; MYSC; 1.			
DR	SMART; SM00326; SH3; 1.			
DR	PROSITE; PS50002; SH3; 1.			
KT	Myosin; ATP-binding; Phosphorylation; Multigene family; SH3 domain.			
FT	DOMAIN	1	670	MYOSIN HEAD-LIKE.
FT	DOMAIN	671	907	TAIL HOMOLOG REGION 1 (TH.1).

```

FT DOMAIN 908 1089 GUY/PRO/ALA-RICH (TH.2.).
FT DOMAIN 1090 1147 SH3.
FT NE_BIND 103 110 ATP (POTENTIAL).
SQ SEQUENCE 1147 AA; 124958 MW; B76DE9C076381054 CRC64;

Query Match
Best Local Similarity 32.4%; Score 49.5; DB 1; Length 1147;
Matches 14; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY 1 YEREKLRVHTQTGGDYFK-FITPLA 25
| | | | | | | | | | | | | | |
Db 179 YLEKSRVTFQTRGERSHFHFYQLLA 204

RESULT 3
TS99 STAU
ID TS99 STAU STANDARD; PRT; 261 AA.
AC P23736;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Type II restriction enzyme Sau96I (EC 3.1.21.4) (Endonuclease Sau96I)
(G.Sau96I).
GN SAU96IR.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxId=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PS96;
RX MEDLINE=90370441; PubMed=2204026;
RA Szilak L., Venetianer P., Kiss A.;
RT "Cloning and nucleotide sequence of the genes coding for the Sau96I
restriction and modification enzymes."
RL Nucleic Acids Res. 18:4659-4664(1990).
CC -1- FUNCTION: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE GGNC AND
CLEAVES AFTER G-1.
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give
specific double-stranded fragments with terminal 5'-phosphates.
CC -1- SIMILARITY: TO R.EC04711.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC CC
CC EMBL; X53096; CAA37259.1; -.
DR DR
DR PIR; S12706; S12706.
DR REBASE; 1607; Sau96I.
KW Hydrolyase; Endonuclease; Nuclease; Restriction system.
SQ SEQUENCE 261 AA; 30486 MW; 75338416FC0D54677 CRC64;

Query Match
Best Local Similarity 32.0%; Score 49; DB 1; Length 261;
Matches 10; Conservative 5; Mismatches 5; Indels 4; Gaps 1;

QY 3 REKLRVTHQ---TTGGDYFKFIT 22
| | | | | | | | | | | | | | |
Db 172 REKPRFSHPVRIRIASGDQFYKIVT 195

RESULT 4
FOLC_CAEEL
ID FOLC_CAEEL STANDARD; PRT; 510 AA.
AC 009509;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Putative folylpolyglutamate synthase (EC 6.3.2.17) [Folylpoly-gamma-
glutamate synthetase] (FPGS).

```


GN F2585.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Tatch A.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CONVERSION OF FOLATES TO POLYGLUTAMATE DERIVATIVES
 CC (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + [tetrahydrofolyl-[glu]](N) + L-glutamate
 CC = ADP + phosphate + [tetrahydrofolyl-[glu]](N+1).
 CC -1- SIMILARITY: BELONGS TO THE POLYGLUTAMATE SYNTHASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U23172; AAC46527.1; -
 CC DR HSSP: P15925; 1FGS.
 DR WormRep: F2585.6; CE01923.
 DR InterPro: IPR001645; Fpolylgl-synthase.
 DR InterPro: IPR0040713; Mur_ligase.
 DR Pfam: PF01225; Mur_ligase_C.
 DR Pfam: PF02875; Mur_ligase_1.
 DR PROSITE: PS01011; FOLYLPOLYGLU_SYNT_1; 1.
 DR PROSITE: PS01012; FOLYLPOLYGLU_SYNT_2; 1.
 KW Hypothetical protein; Ligase; One-carbon metabolism; ATP-binding.
 FT NP_BIND 95 101 ATP (POTENTIAL).
 SQ SEQUENCE 510 AA; 56432 MW; 12A65157DFED84A CRC64;

Query Match 31.7%; Score 48.5; DB 1; Length 510;
 Best Local Similarity 28.3%; Pred. No. 7.5;
 Matches 13; Conservative 5; Mismatches 5; Indels 23; Gaps 1;

OY 3 REKRLVTHQTGDQ-----YKFTLLA 25
 DB 129 REKIQDGPVSEQMAEEFHHYDIKREHSDNMPAYKFFLLA 174

RESULT 5
 MISC_ACACA STANDARD; PRT; 1168 AA.
 AC P10569;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin IC heavy chain.
 GN MIC.
 OS Acanthamoeba castellanii (Amoeba).
 OC Eukaryota; Acanthamoebidae; Acanthamoeba.
 OX NCBI_TaxID=5755;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8016163; PubMed=3477803;
 RA Jung G., Korn E.D., Hammer J.A. III;
 RT "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
 RT and non-myosin-like sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).
 RN [2]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=86259656; PubMed=3014500;
 RA Hammer J.A. III, Jung G., Korn E.D.;
 RT "Genetic evidence that Acanthamoeba myosin I is a true myosin."
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4655-4659(1986).
 RN [3]

RP PHOSPHORYLATION SITE.
 RX MEDLINE=9003074; PubMed=2530230;
 RA Bressan H., Lynch T.J., Martin B., Korn E.D.;
 RT "The localization and sequence of the phosphorylation sites of
 RT Acanthamoeba myosins I. An improved method for locating the
 RT phosphorylated amino acid."
 RL J. Biol. Chem. 264:19340-19348(1989).
 CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
 CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
 CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
 CC -1- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
 CC MYOSINS I CAN THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND VICE
 CC VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER WITH
 CC THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE MOLECULES
 CC OR MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
 CC -1- MISCELLANEOUS: THIS ORGANISM EXPRESSES AT LEAST THREE ISOFORMS OF
 CC MYOSIN I HEAVY-CHAIN, ENCODED BY GENES MIA, MIB, AND MIC.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE MYOSIN IB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J02974; AAA27707.1; -
 CC DR PIR: A33891; MMAXIC.
 DR HSSP: P08799; 1MND.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00018; SH3_1.
 DR Pfam: PF00063; myosin_head_1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PRODOM: PD000066; SH3_1.
 DR PRODOM: PD000355; myosin_head_1.
 DR SMART: SM00242; MySc; 1.
 DR SMART: SM00326; SH3_1.
 DR PROSITE: PS50002; SH3_1.
 DR PROSITE: PS50002; SH3_1.
 KW Myosin: ATP-binding; Phosphorylation; Multigene family; SH3 domain.
 FT DOMAIN 1 670 MYOSIN HEAD-LIKE.
 FT DOMAIN 1 922 TAIL HOMOLOG REGION 1 (TH.1).
 FT DOMAIN 923 975 GLY/PRO/ALA-RICH (TH.2).
 FT DOMAIN 976 1035 SH3.
 FT DOMAIN 1036 1168 GLY/PRO/ALA-RICH (TH.2).
 FT NP_BIND 101 108 ATP (POTENTIAL).
 FT MOD_RES 311 311 PHOSPHORYLATION.
 SQ SEQUENCE 1168 AA; 127309 MW; D07084B373A37A32 CRC64;

Query Match 31.7%; Score 48.5; DB 1; Length 1168;
 Best Local Similarity 50.0%; Pred. No. 19;
 Matches 13; Conservative 3; Mismatches 9; Indels 1; Gaps 1;

OY 1 YERKRLVTHQTGDQYK-FITLLA 25
 DB 175 YLEKSRVYVYQNGERNHFIFQLLA 200

RESULT 6
 PST2_SCHPO STANDARD; PRT; 1075 AA.
 AC O13919;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Paired amphipathic helix protein pst2 (Sin3 homolog).
 GN PST2 OR SPAC23C11.15.
 OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins S., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gattles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbitts J.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grynopre B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs W., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas R., Rochet M., Galland C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shipakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 RN [2]
 RP GENE NAME.
 RX MEDLINE=99147069; PubMed=10022921;
 RA Dang V.D., Benedik M.J., Ekwall K., Choi J., Allshire R.C.,
 RA Levin H.L.;
 RT "A new member of the Sin3 family of corepressors is essential for
 RT cell viability and required for retroelement propagation in fission
 RT yeast."
 RL Mol. Cell. Biol. 19:2351-2365(1999).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: TO YEAST PAIRED AMPHIPATHIC HELIX PROTEIN (SIN3).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z98359; CAB1171.1; -
 DR InterPro: IPR003822; Pfam: PF02671; PAH; 3.
 KW Nuclear protein.
 SO SEQUENCE 1075 AA; 124848 MW; 1AD301DB4EB1AFB CRC64;
 OY 1 YERKLT-----RYTHQTGDQYKFTLLADQHS 29
 Db 845 YTDKLTWSAKOVHHIVSDGKRYFTSLVEQNS 878
 Query Match 30.4%; Score 46.5; DB 1; Length 1075;
 Best Local Similarity 32.4%; Pred. No. 36;
 Matches 11; Conservative 6; Mismatches 12; Indels 5; Gaps 1;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 19.4 kDa protein Y4WG.
 GN Y4WG.
 OS Rhizobium sp. (strain NGR234).
 OC Plasmid sym pNGR234a.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97305956; PubMed=9163424;
 RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
 RA Perret X.;
 RT "Molecular basis of symbiosis between Rhizobium and legumes."
 RL Nature 387:394-401(1997).
 CC -1- SIMILARITY: NONE OBTAINED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000103; AAB91914.1; -
 DR Hypothetical protein; Plasmid.
 KW SEQUENCE 181 AA; 19427 MW; 3F63E57811687B4C CRC64;
 SO SEQUENCE 181 AA; 19427 MW; 3F63E57811687B4C CRC64;
 OY 12 TTGDQYFKFTLLADQ 27
 Db 60 TTADQGLAFVSLIADQ 75
 Query Match 29.4%; Score 45; DB 1; Length 181;
 Best Local Similarity 50.0%; Pred. No. 8;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 RESULT 8
 ID TYTR_TRYB STANDARD; PRT; 492 AA.
 AC P33051;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trypanothione reductase (EC 1.6.4.8) (TR) (N(1),N(8) -
 DE bis(glutathionyl)spermidine reductase).
 GN TPR.
 OS Trypanosoma brucei brucei.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ITFAT 1.1;
 RX MEDLINE=93086418; PubMed=1453951;
 RA Aboagye-Kwarteng T., Smith K., Fairlamb A.H.;
 RT "Molecular characterization of the trypanothione reductase gene from
 RT Crithidia fasciculata and Trypanosoma brucei: comparison with other
 RT flavoprotein disulphide oxidoreductases with respect to substrate
 RT specificity and catalytic mechanism."
 RL Mol. Microbiol. 6:3089-3099(1992).
 CC -1- FUNCTION: TRYpanOTHIONE IS THE PARASITE ANALOG OF GLUTATHIONE;
 CC THIS ENZYME IS THE EQUIVALENT OF GLUTATHIONE REDUCTASE.
 CC -1- CATALYTIC ACTIVITY: NADPH + trypanothione = NADP(+) + reduced
 CC trypanothione.
 CC -1- COFACTOR: FAD.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
 CC OXIDOREDUCTASES CLASS-I.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63188; CAA44870.1; -.
DR PIR; S28003; S28003.
DR HSSP; P28593; INDA.
DR InterPro: IPR001327; FAD_pyr_redox.
DR InterPro: IPR001100; Pyr_redox.
DR InterPro: IPR001864; tryptph_redctse.
DR InterPro: IPR004099; pyr_redox_dlm.
DR Pfam; PF00070; pyr_redox; 1.
DR Pfam; PF02852; pyr_redox_dlm; 1.
DR PRINTS; PR00411; PNEDTASEI.
DR PRINTS; PR00470; TRYPARNDTASE.
DR PRODOM; PD000139; FAD_pyr_redox; 1.
DR PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
KW Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP.
FT NP_BIND 6 52 57 FAD (ADP PART) (PROBABLE).
FT DISULFID 52 57 REDOX-ACTIVE.
FT NP_BIND 317 327 FAD (FLAVIN PART) (BY SIMILARITY).
FT ACT_SITE 461 461 BY SIMILARITY.
SQ SEQUENCE 492 AA; 53284 MW; FFCAP2FL1B66TFPF4 CRC64;

Query Match
Best Local Similarity 29.4%; Score 45; DB 1; Length 492;
Matches 11; Conservative 5; Mismatches 12; Indels 8; Gaps 1;

OY 2 EREKLV-----THQTTGQDYFKFTLLADQHS 29
DB 385 EFEKVAVYMSFPTLMHNISGRYKKFKVAKIYNHS 420

RESULT 9
VARTL_SULSO
ID VARTL_SULSO STANDARD: PRT; 701 AA.
AC Q9UWV3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE V-type ATP synthase subunit I (EC 3.6.3.14) (V-type ATPase subunit I).
GN ATP1 OR S500539 OR C21_040.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=20165948; PubMed=10701121;
RA Charlebois R.L., Singh R.K., Chan-Welher C.C.-Y., Allard G., Chow C.,
RA Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,
RA Kushwaha N., Lafleur E., Medina N., Peng X., Penny S.L., She Q.,
RA St. Jean A., Sengen C.W., Young F., Zivanovic Y., Doolittle W.F.,
RA Ragan M.A., van der Oost J.
RT "Gene content and organization of a 281-kbp contig from the genome of
RT the extremely thermophilic archaeon, Sulfolobus solfataricus P2."
RL Genome 43:116-136(2000).

RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thl-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,

```

```

RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y18930; CAB57740.1; -.
DR EMBL; AE006686; AAK40876.1; -.
DR InterPro: IPR002490; V_ATPase_sub116.
DR Pfam; PF01496; V_ATPase_sub_a; 1.
KW Hydrolyase; Hydrogen ion transport; Transmembrane; Complete proteome.
SQ SEQUENCE 701 AA; 80131 MW; FBC53187B01B31EA CRC64;

Query Match
Best Local Similarity 29.4%; Score 45; DB 1; Length 701;
Matches 8; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

OY 2 EREKLVTHQTTGQDYFKFTLLADQ 27
DB 154 DEDKRLRLNQTIGDSNFYYTRFGE 179

RESULT 10
MOKD_SCHPO
ID MOKD_SCHPO STANDARD: PRT; 2358 AA.
AC Q9Y719; Q94638;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell wall alpha-1,3-glucan synthase mok13 (EC 2.4.1.183).
GN MOK13 OR SPC16D10.05.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Katayama S., Dai H., Arellano M., Perez P., Toda T.;
RT "Fission yeast alpha-glucan synthase MOK1 localizes closely with actin
RT and play a role essential for cell morphogenesis and protein kinase C
RT function."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
RC STRAIN-972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voiclaert G., Aert R., Robben J., Grymprez B.,

```

```

RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Filz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger F., Zimmermann W., Wedler H., Wambutt R., Parnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Belito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombe W.R., Paulsen I., Potashkin J.,
RA Sipkavski G.V., Ussery D., Barrel B.G., Nurse P.;
RT "The genome sequence of Schistosoma mansoni pombe.";
RL Nature 415:871-880(2002).
CC -1- CATALYTIC ACTIVITY: UDP-glucose + {alpha-D-glucosyl-(1,3)}(N) =
CC UDP + {alpha-D-glucosyl-(1,3)}(N+1).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB018382; BAB559.1; -.
DR EMBL; AL035637; CAB38509.1; -.
DR InterPro: IPR000461; Alpha_amyrase.
DR Pfam: PF00128; alpha-amyrase.1.
KW Cell wall; transferase; glycosyltransferase.
FT CONFLICT 120 120 V->VRVWLCLSTNKV (IN REF. 1).
SQ SEQUENCE 2358 AA; 269192 MW; A175577C9D8AD731 CRC64;

Query Match 29.4%; Score 45; DR 1; Length 2358;
Best Local Similarity 42.1%; Pred. No. 1.5e+02;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 EREKLRVHTQTDGYEKF 20
Dbb 858 ENKLYINHAKPGADLRF 876

RESULT 11
RYRL_PIG STANDARD; PRT; 5035 AA.
AC P16960;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ryanodine receptor 1 (Skeletal muscle-type ryanodine receptor) (RYR1)
DE (RYR-1) (Skeletal muscle calcium release channel).
GN RYR1 OR CRC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Norwegian Landrace; TISSUE-Skeletal muscle;
RX MEDLINE=93036581; PubMed=1329581;
RA Harbitz I., Kristensen T., Bosnes M., Kran S., Davies W.;
RT "DNA sequence of the skeletal muscle calcium release channel cDNA and
RT verification of the Arg615->Cys615 mutation, associated with porcine
RT malignant hyperthermia, in Norwegian landrace pigs.";
RL Anim. Genet. 23:395-402(1992).
RN [2]
RP SEQUENCE OF 1129-2801 FROM N.A.
RA Brenig B.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1129-2643 FROM N.A.
RC STRAIN-German Landrace; TISSUE-Liver;
RX MEDLINE=94117003; PubMed=8288238;

```

```

RA Leeb T., Schmolzl S., Brem G., Brenig B.;
RT "Genomic organization of the porcine skeletal muscle ryanodine
RT receptor (RYR1) gene coding region 4624 to 7929.";
RL Genomics 18:349-354(1993).
RN [4]
RP SEQUENCE OF 4785-5035 FROM N.A.
RX MEDLINE=91065640; PubMed=2174405;
RA Harbitz I., Chowdhary B., Thomsen P.D., Davies W., Kaufman U.,
RA Kran S., Gustavsson I., Christensen K., Haug J.G.;
RT "Assignment of the porcine calcium release channel gene, a candidate
RT for the malignant hyperthermia locus, to the 6p11-->q21 segment of
RT chromosome 6.";
RL Genomics 8:243-248(1990).
CC -1- FUNCTION: COMMUNICATION BETWEEN TRANSVERSE-TUBULES AND
CC SARCOPLASMIC RETICULUM. CONTRACTION OF SKELETAL MUSCLE IS
CC TRIGGERED BY RELEASE OF CALCIUM IONS FROM SR FOLLOWING
CC DEPOLARIZATION OF T-TUBULES.
CC -1- SUBUNIT: HOMOTETRAMER (POTENTIAL).
CC -1- MISCELLANEOUS: The calcium release channel is modulated by calcium
CC ions, magnesium ions, ATP and calmodulin.
CC -1- MISCELLANEOUS: The calcium release channel activity resides in the
CC C-terminal region while the remaining part of the protein
CC constitutes the 'foot' structure spanning the junctional gap
CC between the SR and the T-tubule. It is possible that the foot
CC structure interacts with the cytoplasmic region of the
CC dihydropyridine receptor.
CC -1- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-
CC release channel in junctional SR and modulates its activity.
CC -1- SIMILARITY: BELONGS TO THE RYANODINE RECEPTOR FAMILY.
CC -1- SIMILARITY: CONTAINS 3 SPRY DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62880; CAA44674.1; ALT_SEQ.
DR EMBL; X68247; CAA48318.1; -.
DR EMBL; X69465; CAA49225.1; -.
DR EMBL; M32501; AAA31022.1; -.
DR PIR; A37105; A37105.
DR InterPro: IPR000699; Ca-rel_channel.
DR InterPro: IPR001682; Ca/Na_pore.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000636; M-channel_nlg.
DR InterPro: IPR003608; MIR.
DR InterPro: IPR003032; RYR.
DR InterPro: IPR001215; Ryanodn_receptor.
DR InterPro: IPR003878; SPRY_domain.
DR InterPro: IPR003877; SPRY_receptor.
DR Pfam; PF00520; Ion_Trans; 1.
DR Pfam; PF00622; SPRY; 3.
DR Pfam; PF01365; RYR-ITPR; 2.
DR Pfam; PF02026; RYR; 4.
DR Pfam; PF02815; MIR; 4.
DR PRINTS: PR00795; RYANODINER.
DR SMART; SM00472; MIR; 4.
DR SMART; SM00449; SPRY; 3.
DR Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat;
KW Phosphorylation; Glycoprotein.
FT DOMAIN 1 3124
FT TRANSMEM 3125 3145 CYTOPLASMIC.
FT TRANSMEM 3189 3207 M' (POTENTIAL).
FT TRANSMEM 3189 3207 M' (POTENTIAL).
FT TRANSMEM 3980 3999 M1 (POTENTIAL).
FT TRANSMEM 4018 4036 M2 (POTENTIAL).
FT TRANSMEM 4274 4297 M3 (POTENTIAL).
FT TRANSMEM 4339 4359 M4 (POTENTIAL).
FT TRANSMEM 4557 4578 M5 (POTENTIAL).
FT TRANSMEM 4646 4669 M6 (POTENTIAL).
FT TRANSMEM 4787 4807 M7 (POTENTIAL).

```

```

FT TRANSMEM 4835 4854 M8 (POTENTIAL).
FT TRANSMEM 4877 4896 M9 (POTENTIAL).
FT TRANSMEM 4921 4935 M10 (POTENTIAL).
FT DOMAIN 659 797 SPRY 1.
FT DOMAIN 1085 1208 SPRY 2.
FT DOMAIN 1430 1570 SPRY 3.
FT DOMAIN 1874 1925 GLU-RICH (ACIDIC).
FT DOMAIN 4458 4526 PRO-RICH.
FT DOMAIN 842 2960 6 X APPROXIMATE REPEATS.
FT REPEAT 842 955 1.
FT REPEAT 956 1069 2.
FT REPEAT 1345 1360 3 (INCOMPLETE).
FT REPEAT 1373 1388 4 (INCOMPLETE).
FT REPEAT 2272 2286 5.
FT REPEAT 2847 2860 6.
FT MOD_RES 2844 2844 PHOSPHORYLATION (BY PKA AND PKC) (BY SIMILARITY).
FT MOD_RES 3947 3947 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 4320 4320 PHOSPHORYLATION (POTENTIAL).
FT CARBOHYD 3467 3467 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3475 3475 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3904 3904 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3945 3945 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4144 4144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4862 4862 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 2092 2092 A -> P (IN REF. 2 AND 3).
SQ SEQUENCE 5035 AA; 565317 MW; E00613F2027B94A4 CRC64;

```

```

Query Match          29.4%; Score 45; DB 1; Length 5035;
Best Local Similarity 37.5%; Pred. No. 3.7e+02;
Matches 9; Conservative 6; Mismatches 5; Indels 4; Gaps 1;

```

```

QY 4 EKIRVTHQTTGDDYFKFTLLDQ 27
   : ||: | ||| : : : | : |
Db 278 OPLRIHVTGTG----RYLALIEDQ 297

```

```

RESULT 12
PEX3_PICPA
ID PEX3_PICPA STANDARD; PRT; 455 AA.
AC Q92262;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Peroxisomal membrane protein PMS2 (Peroxin-3).
GN PEX3 OR PMS2.
OS Pichia pastoris (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4922;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97115764; PubMed=8955066;
RA Subramani S.;
RT "Protein translocation into peroxisomes.";
RL J. Biol. Chem. 271:32483-32486(1996).
CC -1- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL_272390; CAA96530.1; -
KW Transmembrane; Peroxisome.
FT DOMAIN 1 15 MATRIX (POTENTIAL).
FT TRANSMEM 16 33 POTENTIAL.
FT DOMAIN 34 455 CYTOPLASMIC (POTENTIAL).
FT SEQUENCE 455 AA; 51973 MW; 6853C5B5A5C67BC34 CRC64;

```

```

Query Match          29.1%; Score 44.5; DB 1; Length 455;
Best Local Similarity 40.0%; Pred. No. 28;
Matches 10; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

```

```

QY 1 YEREKL-RVTHQTTGDDYFKFTLL 24
   : : : | ||| | | : : |
Db 49 FAKQIKRFRHQTOSDCYMTFLSL 73

```

```

RESULT 13
ALBU_PIG
ID ALBU_PIG STANDARD; PRT; 605 AA.
AC P08835; Q29018;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Serum albumin precursor (Fragment).
GN ALB.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89016582; PubMed=3174440;
RA Baldwin G.S., Weinstein J.;
RT "Nucleotide sequence of porcine liver albumin.";
RL Nucleic Acids Res. 16:9045-9045(1988).
CC -1- FUNCTION: SERUM ALBUMIN, THE MAIN PROTEIN OF PLASMA, HAS A GOOD
CC BINDING CAPACITY FOR WATER, CA++, NA+, K+, FATTY ACIDS, HORMONES,
CC BILIRUBIN AND DRUGS. ITS MAIN FUNCTION IS THE REGULATION OF THE
CC COLLOIDAL OSMOTIC PRESSURE OF BLOOD.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PLASMA.
CC -1- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
CC -1- SIMILARITY: CONTAINS 3 ALBUMIN DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL_X12422; CAA30970.1; -
DR EMBL_M36787; AAA30988.1; -
DR PIR_S01382; ABGCS.
DR HSSP_P02768; IE7H.
DR InterPro; IPR000264; Serum_albumin.
DR Pfam; PF00273; transport_prct; 3.
DR ProDom; PD002486; Serum_albumin; 1.
DR SMART; SM00103; ALBUMIN; 3.
DR PROSITE; PS00212; ALBUMIN; 3.
KW Plasma; Metal-binding; Lipid-binding; Albumin; Repeat; Signal;
KW Copper.
FT NON_TER 1 1
FT SIGNAL <1 16 BY SIMILARITY.
FT PROPEP 17 22 BY SIMILARITY.
FT CHAIN 23 605 SERUM ALBUMIN.
FT DOMAIN 23 202 ALBUMIN 1.
FT DOMAIN 209 394 ALBUMIN 2.
FT DOMAIN 401 592 ALBUMIN 3.
FT METAL 31 31 COPPER (BY SIMILARITY).
FT METAL 75 84 BY SIMILARITY.
FT DISULFD 97 113 BY SIMILARITY.
FT DISULFD 112 123 BY SIMILARITY.
FT DISULFD 145 190 BY SIMILARITY.
FT DISULFD 189 198 BY SIMILARITY.
FT DISULFD 221 267 BY SIMILARITY.
FT DISULFD 266 274 BY SIMILARITY.

```

```

FT DISULFID 286 300 BY SIMILARITY.
FT DISULFID 299 310 BY SIMILARITY.
FT DISULFID 337 382 BY SIMILARITY.
FT DISULFID 381 390 BY SIMILARITY.
FT DISULFID 413 459 BY SIMILARITY.
FT DISULFID 458 469 BY SIMILARITY.
FT DISULFID 482 498 BY SIMILARITY.
FT DISULFID 497 508 BY SIMILARITY.
FT DISULFID 535 580 BY SIMILARITY.
FT DISULFID 579 588 BY SIMILARITY.
FT CONFLICT 562 562 E -> D (IN REF. 1; AAA0998).
SQ SEQUENCE 605 AA; 69410 MW; 3E556BDDDA1FAF6 CRC64;

Query Match 29.1%; Score 44.5; DB 1; Length 605;
Best Local Similarity 40.7%; Pred. No. 38;
Matches 11; Conservative 5; Mismatches 8; Indels 3; Gaps 2;

QY 5 KLRVTH--QTGGDYKFKFTLLA-DOH 28
   1 : 1 : 1:1111 : 1:1 11
DB 26 KSEIAHRKDLGEQYKFKGLVLAFSOH 52

RESULT 14
Y032_MYCGE STANDARD; PRT; 666 AA.
AC P47278;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG032.
GN MG032.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischman R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
CC -1- SIMILARITY: BELONGS TO THE MG032 / MG036 / MG288 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39692; AAC71248.1; -
DR TIGR: MG032; -
DR InterPro: IPR004306; MG032/096/288_1.
DR InterPro: IPR004319; MG032/096/288_2.
DR Pfam: PF03072; DUF237; 1.
DR Pfam: PF03086; DUF240; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 666 AA; 77302 MW; D703C107A8E2AB7F CRC64;

Query Match 29.1%; Score 44.5; DB 1; Length 666;
Best Local Similarity 52.6%; Pred. No. 43;
Matches 10; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 5 KLRVTHQTGGD-QYKFKFT 22
   111111: 11 1:: 1
DB 576 KLRVTHQPDGDPYRYRLHT 594

```

```

RESULT 15
SCAA_HUMAN STANDARD; PRT; 669 AA.
ID SCAA_HUMAN
AC P37088;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Amiloride-sensitive sodium channel alpha-subunit (Epithelial Na+
DE channel alpha subunit) (Alpha ENaC) (Nonvoltage-gated sodium channel
DE 1 alpha subunit) (SCN5A) (Alpha NaCh).
GN SCN5A OR SCN5L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=94105144; PubMed=8278374;
RA Volley N., Lingueglia E., Champigny G., Mattei M.-G., Walmann R.,
RA Lazdunski M., Barbry P.;
RT "The lung amiloride-sensitive Na+ channel: biophysical properties,
RT pharmacology, ontogenesis, and molecular cloning.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:247-251(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94295729; PubMed=8023962;
RA McDonald F.J., Snyder P.M., McCray P.B., Welsh M.J.;
RT "Cloning, expression, and tissue distribution of a human amiloride-
RT sensitive Na+ channel.";
RL Am. J. Physiol. 266:L728-L734(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=98316780; PubMed=9654208;
RA Ludwig M., Bolkenius U., Wickerl L., Marynen P., Bidlingmaier F.;
RT "Structural organisation of the gene encoding the alpha-subunit of
RT the human amiloride-sensitive epithelial sodium channel.";
RL Hum. Genet. 102:576-581(1998).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=99374783; PubMed=10447117;
RA Chow Y.H., Wang Y., Plumb J., O'Brodovich H., Hu J.;
RA "Hormonal regulation and genomic organization of the human amiloride-
RT sensitive epithelial sodium channel alpha subunit gene.";
RL Pediatr. Res. 46:208-214(1999).
CC -1- FUNCTION: SODIUM PERMEABLE NON-VOLTAGE-SENSITIVE ION CHANNEL
CC INHIBITED BY THE DIURETIC AMILORIDE. MEDIATE THE ELECTRODIFFUSION
CC OF THE LUMINAL SODIUM (AND WATER, WHICH FOLLOWS OSMOTICALLY)
CC THROUGH THE APICAL MEMBRANE OF EPITHELIAL CELLS. CONTROLS THE
CC REABSORPTION OF SODIUM IN KIDNEY, COLON, LUNG AND SWEAT GLANDS.
CC ALSO PLAYS A ROLE IN TASTE PERCEPTION.
CC -1- SUBUNIT: HETEROTETRAMER OF TWO ALPHA, ONE BETA AND ONE GAMMA
CC SUBUNIT. A DELTA SUBUNIT CAN REPLACE THE ALPHA SUBUNIT.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- DISEASE: DEFECTS IN SCN5A ARE ONE OF THE CAUSES OF
CC PSEUDOHYPONATREMISM TYPE 1 (PHA1). A RARE SALT WASTING DISEASE
CC CHARACTERIZED BY AN OFTEN FAMILIAR PRESENTATION IN THE NEONATAL
CC PERIOD WITH DEHYDRATION, HYONATRAEMIA, HYPERKALAEMIA, METABOLIC
CC ACIDOSIS, FAILURE TO THRIVE AND WEIGHT LOSS.
CC -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNEL
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X76180; CAA53773.1; -

```

DR EMBL: L29007; AAA21813.1; -;
 DR EMBL: 292978; CAB07505.1; -;
 DR EMBL: 292979; CAB07505.1; JOINED.
 DR EMBL: 292980; CAB07505.1; JOINED.
 DR EMBL: 292981; CAB07505.1; JOINED.
 DR EMBL: AF060913; AAD28355.1; -;
 DR EMBL: AF060910; AAD28355.1; JOINED.
 DR EMBL: AF060911; AAD28355.1; JOINED.
 DR EMBL: AF060912; AAD28355.1; JOINED.
 DR GeneW; HGNC:10599; SCN1A.
 DR MIM: 600228; -;
 DR MIM: 17735; -;
 DR MIM: 264350; -;
 DR InterPro; IPR004724; Enac.
 DR InterPro; IPR001873; Na+channel_ASC.
 DR Pfam; PF00858; ASC; 1.
 DR PRINTS; PR01078; AMINACHANNEL.
 DR TIGRFAMs; TIGR00859; ENAC; 1.
 DR PROSITE; PS01206; ASC; 1.
 KW Ionic channel; Transmembrane; Ion transport; Glycoprotein.
 FT DOMAIN 1 85 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 86 106 POTENTIAL.
 FT DOMAIN 107 562 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 563 583 POTENTIAL.
 FT DOMAIN 584 669 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 669 AA: 75703 MW: 20CF342E7DF32E72 CRC64;

Query Match 29.1%; Score 44.5; DB 1; Length 669;
 Best Local Similarity 45.2%; Pred. No. 43;
 Matches 14; Conservative 4; Mismatches 10; Indels 3; Gaps 2;

QY 2 EREKL-RVTHQTGD--QYKFTITLADQHS 29
 DB 146 ELEELDRITEQTLPDIKYSSFTTLVAGSRS 176

Search completed: January 29, 2003, 06:46:42
 Job time : 21.8246 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 06:44:49 ; Search time 64.1053 Seconds
(without alignments)
93.212 Million cell updates/sec

Title: US-09-627-165D-19
Perfect score: 153
Sequence: IYERKLRVTHQTGDQYKFFITLADQHS 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:REMBL_21:*
2: SP:archaea:*
3: SP:bacteria:*
4: SP:fungi:*
5: SP:human:*
6: SP:mammal:*
7: SP:mnc:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:rodent:*
12: SP:virus:*
13: SP:vertebrate:*
14: SP:unclassified:*
15: SP:rivirus:*
16: SP:bacteriophage:*
17: SP:archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	114	74.5	249	10 Q8RXH7	Q8RXH7 viscum albu
2	114	74.5	565	10 Q8W243	Q8W243 viscum albu
3	110	71.9	29	10 Q957D0	Q957D0 viscum albu
4	110	71.9	531	10 Q8RXH6	Q8RXH6 viscum albu
5	55.5	36.3	848	12 Q8QLJ2	Q8QLJ2 mamestra co
6	55	35.9	117	17 Q9HUK9	Q9HUK9 thernoplasm
7	51.5	33.7	454	16 Q55932	Q55932 synechocyst
8	51	33.3	464	5 Q18854	Q18854 caenorhabdi
9	50	32.7	116	17 Q979N3	Q979N3 thernoplasm
10	50	32.7	161	16 Q8ZSA4	Q8ZSA4 anabaena sp
11	48.5	31.7	521	5 Q95002	Q95002 acanthamoeb
12	48.5	31.7	1186	5 Q61080	Q61080 agrobacteri
13	48	31.4	50	16 Q8U5B5	Q8U5B5 archaeglob
14	48	31.4	190	17 Q29564	Q29564 archaeglob
15	48	31.4	318	16 Q8VM49	Q8VM49 anabaena sp
16	47	30.7	654	11 Q9D0H7	Q9D0H7 mus musculu

17	47	30.7	809	12 Q8QNB5	Q8QNB5 ectocarpus
18	47	30.7	2556	5 Q9VXP7	Q9VXP7 dirosophila
19	46.5	30.4	421	2 Q54496	Q54496 streptococc
20	46.5	30.4	841	17 Q59612	Q59612 pyrodictus
21	46	30.1	80	10 Q9XE28	Q9XE28 pinus strob
22	46	30.1	278	10 Q9SXX1	Q9SXX1 oryza sativ
23	46	30.1	367	10 Q9SNG2	Q9SNG2 oryza sativ
24	46	30.1	381	16 Q8X9W2	Q8X9W2 escherichia
25	46	30.1	544	16 Q8G4I8	Q8G4I8 rhizobium l
26	46	30.1	571	10 Q8W040	Q8W040 arabidopsis
27	46	30.1	689	2 Q50642	Q50642 porphyromon
28	46	30.1	700	17 Q26460	Q26460 methanobact
29	46	30.1	1036	5 Q17382	Q17382 caenorhabdi
30	46	30.1	1100	5 Q19901	Q19901 caenorhabdi
31	45.5	29.7	58	16 Q9ZFI6	Q9ZFI6 rhizobium l
32	45.5	29.7	226	2 Q9ZFI6	Q9ZFI6 streptomyce
33	45.5	29.7	1326	5 Q9W105	Q9W105 dirosophila
34	45	29.4	88	16 Q9PFN2	Q9PFN2 xyloella fas
35	45	29.4	140	16 Q97LT8	Q97LT8 clostridium
36	45	29.4	174	12 Q84576	Q84576 parametium
37	45	29.4	362	5 Q960V7	Q960V7 dirosophila
38	45	29.4	383	16 Q9X853	Q9X853 streptomyce
39	45	29.4	399	5 Q20459	Q20459 caenorhabdi
40	45	29.4	488	16 Q8Z622	Q8Z622 salmoneilla
41	45	29.4	587	5 Q8T0H8	Q8T0H8 dirosophila
42	45	29.4	620	5 Q21374	Q21374 caenorhabdi
43	45	29.4	717	9 Q8SD07	Q8SD07 pseudomonas
44	45	29.4	758	5 Q9V9X1	Q9V9X1 dirosophila
45	45	29.4	1128	6 P79271	P79271 sus scrofa

ALIGNMENTS

RESULT 1

ID	Q8RXH7	PRELIMINARY:	PRT:	249 AA.
AC	Q8RXH7:			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Lectin chain A isoform 2 (Fragment).			
OS	Viscum album (European mistletoe).			
OC	Eukaryota: Viridiplantae: Streptophyta: Tracheophyta:			
OC	Spermatophyta: Magnoliophyta: eudicotyledons; core eudicots;			
OC	Santalales: Viscaceae: Viscum.			
OX	NCBI_TaxID=3972;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LEAF.			
RA	Parmanaslam M., Stinvaasan A., Singh T.P.;			
RT	"Viscum Album (Indian) mRNA for Mistletoe lectin Chain A, isoform 2."			
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AY081148; AAL87005.1; -			
FT	NON_TER	1	1	
FT	NON_TER	249	249	
SO	SEQUENCE	249 AA;	27944 MW;	89FAFB78309A83B3 CRC64;

Query Match 74.5%; Score 114; DB 10; Length 249;
Best Local Similarity 80.8%; Pred. No. 1.2e-09;
Matches 21; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY	1 YERKLRVTHQTGDQYKFFITLAD 26	1 YERKLRVTHQTGDQYKFFITLAD 26
DB	1 YERKLRVTHQTGDQYKFFITLAD 26	
RESULT 2		
Q8W243	PRELIMINARY:	PRT: 565 AA.
Q8W243:		
DT	01-MAR-2002 (TREMBLrel. 20, Created)	
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)	

DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	VCA precursor.		
OS	Viscum album subsp. coloratum.		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
OC	Samiales; Viscaceae; Viscum.		
OX	NCBI_TaxID=159976;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Park W.-B., Lyu S.;		
RT	"Cloning of Viscum album subsp. coloratum (Korean mistletoe).",		
RL	Biochem. Biophys. Res. Commun. 0:0-0(2002).		
DR	EMBL; AF369961; AAI40417.1;		
DR	InterPro: IPR001899; Gram_pos_anchor.		
DR	InterPro: IPR000772; Ricin_B_lectin.		
DR	InterPro: IPR001574; RIP.		
DR	Pfam; PF00652; Ricin_B_lectin; 6.		
DR	Pfam; PF00161; RIP; 1.		
DR	PRINTS; PR00396; SHIGARICIN.		
DR	SMART; SM00458; RICIN. 2.		
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.		
DR	PROSITE; PS50231; RICIN_B_LECTIN; 2.		
KW	Signal.		
FT	SIGNAL. 1 22	POTENTIAL.	
FT	CHAIN 23 273	VCA ALPHA CHAIN.	
FT	CHAIN 309 565	VCA BETA CHAIN.	
SO	SEQUENCE 565 AA; 62401 MW; 991E3994DA005F11 CRC64;		

```
Query Match:          74.5%; Score 114; DB 10; Length 565;
Best Local Similarity 80.8%; Pred. NO. 3e-09;
Matches    21; Conservative   3; Mismatches     2; Indels      0; Gaps      0;
```

QY 1 YEREKLRVTHQTGGDYERFEITLLAD 26
 ||| :||| :||| :||| :||| |
Db 34 YEKRLRVRVTHQTGGDYERFEITLLRD 59

RESULT 3
Q9S7D0
ID Q9S7D0 PRELIMINARY; PRT: 29 AA.
AC Q9S7D0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
DE MLI A chain toxic lectin (Fragment).
OS Viscum album (European mistletoe).
OC Euxaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Saxifragales; Viscaceae; Viscum.
OX NCBI_TaxID=3972;
RN [1]
RP SEQUENCE.
RX MEDLINE=93081837; PubMed=145045;
RA Dietrich J.B., Ribereau-Gayon G., Jung M.L., Franz H., Beck J.P.,
RA Anton R.;
RT "identity of the N-terminal sequences of the three A chains of
RT mistletoe (Viscum album L.) lectins: homology with rich-
RT toxins and single-chain ribosome-inhibiting proteins.";
RL Anti-cancer drugs 3:507-511(1997).
SQ SEQUENCE 29 AA; 3633 MW; A7C20F764546B2F5 CRC64;

Query Match	71.98%	Score 110;	DB 10;	length 29;
Best Local Similarity	76.98%	Pred. No. 4,9e-10;		
Matches	20; Conservative	4; Mismatches	2;	Indels 0; Gaps 0;
Qy	1 YERERLRVTHQTGGDYEFERITLLAD	26		
	: : :			
Db	1 YERLRVTHQTGGDYEFERITLLAD	26		

RESULT 4	
Q8RXH6	
ID	PRELIMINARY;
Q8RXH6	PRT; 531 AA.

```

AC      08RXH6;
DT      01-JUN-2002 (TREMBlrel. 21, Created)
DT      01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE      01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DR      Lectin chain A isoform 1 (Fragment).
OS      Viscum album (European mistletoe).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots;
OC      Santalales; Viscaceae; Viscum.
OX      NCBI_TaxID=3972;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      TISSUE=LEAF;
RA      Parmasivam M., Misra V., Srinivasan A., Singh T.P.;
RT      "Viscum album (Indian) mrwa for Mistleoe lectin chain A isoform 1 and
RL      chain B.";
RT      Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AY081149; AL87006.1; -.
FT      NON_TER      1
FT      CHAIN      1
FT      CHAIN      1      249      LECTIN CHAIN A ISOFORM 1.
FT      CHAIN      269      531      LECTIN B CHAIN
SQ      SEQUENCE      531 AA; 58802 MW; 18244BEEFEF35422 CRC64;

Query Match
Best Local Similarity 71.9%; Score 110; DB 10; Length 531;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

1 YEREKLVHTQHTQGDGVKFRITLLAD 26
|||:|||||||:|:|||||
1 YERLRKLVHTQHTQGEFRITLLAND 26

```

	RESULT	5
OBOLJ2		
ID	OBOLJ2	PRELIMINARY;
DP	OBOLJ2;	PRT; 848 AA.
DT	01-JUN-2002 (TrEMBLrel. 21, Created)	
DI	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
OS	Hypothetical 99.1 kDa protein.	
DS	Mamestra configurata nucleopolyhedrovirus.	
OC	Viruses; dsDNA viruses, no RNA stage; Baculoviridae;	
NCBI_Taxid=191492;		
NX	[1]	
RN	SEQUENCE FROM N.A.	
RP	STRAIN=90/2;	
RC	MEDLINE=97163493; PubMed=9010313;	
RA	Ll S., Ertlandson M., Moody B., Gillott C.;	
RT	"A physical map of the Mamestra configurata nucleopolyhedrovirus	
RL	genome and sequence analysis of the polyhedrin gene.";	
FT	J. Gen. Virol. 78:265-271(1997).	

RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RX MEDLINE=21884635; Pubmed=11886270;
RA Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., Erlandson M.;
RT "Sequence and Organization of the Mamestra configurata
RL Nucleopolyhedrovirus Genome."; *Virology* 294:106-121(2002).
RM [3]
RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RA Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., Erlandson M.A.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: U59461; AAM09150.1; -
KW Hypothetical protein.
Q SEQUENCE 848 AA: 99132 MW: 73290556/600300 00007

Query Match	36.3%	Score 55.5;	DB 12;	Length 848;
Best Local Similarity	44.0%;	Pred. No. 6.3;		
Matches 11; Conservative	5;	Mismatches 6;	Indels 3;	Gaps 1

```

OY      8 VTHQ---TTGDQYFKFTLLADDS 29
      :||| : | | : | | : | |
Db      98 ITHQMRLASDDMYKVASFLANQHS 122

RESULT 6
O9HJK9 ID O9HJK9 PRELIMINARY: PRT: 117 AA.
AC O9HJK9:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE Hypothetical protein Ta0958.
GN TA0958.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OC NCBI_Taxid=2303;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RX Ruepp A., Granel W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
  Mewes H.-W., Friesman D., Stocker S., Lupas A.N., Baumeister W.;
  RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
  acidophilum."
  RL Nature 407:508-513(2000).
  DR EMBL: AL445066; CAC12087.1;
  KM Hypothetical protein; Complete proteome.
SQ SHQUENCE 117 AA; 13903 MW; C5E17BE5D3E74C41 CRC64;

Query Match 35.9%; Score 55; DB 17; Length 117;
Best Local Similarity 42.9%; Pred. No. 0.87;
Matches 12; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

OY      3 REKLRVTHQTGDQ---YFKFTLLAD 26
      :| : | : | : | : | : | : |
Db      30 RKLELRKSTTGDSRDYPRYEFISILD 57

RESULT 7
O55932 ID O55932 PRELIMINARY: PRT: 454 AA.
AC O55932:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Sensory transduction histidine kinase.
GN SLU0798.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OC NCBI_Taxid=1148;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8530279;
RX Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
  Sugita M., Tabata S.;
  RT "Sequence analysis of the genome of the unicellular cyanobacterium
  Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
  region from map positions 64k to 92k of the genome."
  RL DNA Res. 2:153-166(1995).
  RN (2)
  RP SEQUENCE FROM N.A.
  RX MEDLINE=97061201; PubMed=8905231;
  RX Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
  Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,
  Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
  Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
  Tabata S.;
  RT "Sequence analysis of the genome of the unicellular cyanobacterium
  Synechocystis sp. strain PCC6803. II. Sequence determination of the
  entire genome and assignment of potential protein-coding regions."
  RL DNA Res. 3:109-136(1996).

```

```

CC      -I- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
DR EMBL: D64005; BA01697.1;
DR Interpro: IPR003594; Arpbind_ATPase.
DR Interpro: IPR004358; Bact_sens_pr_C.
DR Interpro: IPR000515; BPD_transp.
DR Interpro: IPR003661; His_kinA.
DR Interpro: IPR004359; HIS_KIN_slg.
DR Pfam: PF02518; HATPase_C; 1.
DR Pfam: PF00512; signal; 1.
DR PRINTS: PR00344; BCTRLSENSOR.
DR SMART: SM00387; HATPase_C; 1.
DR SMART: SM00388; HISKA; 1.
DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_1.
KW Kinase; Phosphorylation; Sensory transduction; Transferase;
KW Complete proteome.
SQ SEQUENCE 454 AA; 50465 MW; FOAB55B5D561D1EE CRC64;

Query Match 33.7%; Score 51.5; DB 16; Length 454;
Best Local Similarity 37.9%; Pred. No. 13;
Matches 11; Conservative 6; Mismatches 9; Indels 3; Gaps 1;

OY      1 YEREKLRVTHQTGDQYFKFTLLADDS 29
      :| : | : | : | : | : | : |
Db      134 FNOETWQTHPTGDRYRQFTLL--HS 159

RESULT 8
O18854 ID O18854 PRELIMINARY: PRT: 464 AA.
AC O18854:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CS4G4.9 protein.
GN CS4G4.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_Taxid=6239;
RN (1)
RP SEQUENCE FROM N.A.
RX Wilkison J.;
  RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
  RN (2)
  RP SEQUENCE FROM N.A.
  RX MEDLINE=99069613; PubMed=9851916;
  RA none;
  RT "Genome sequence of the nematode C.elegans: A platform for
  RT investigating biology."
  RL Science 282:2012-2018(1998).
  DR EMBL: Z75533; CAA99821.1;
  SQ SEQUENCE 464 AA; 54140 MW; 07582ED06C1FC45B CRC64;

Query Match 33.3%; Score 51; DB 5; Length 464;
Best Local Similarity 42.9%; Pred. No. 17;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY      2 EREKLRVTHQTGDQYFKFT 22
      :| : | : | : | : | : | : |
Db      427 EEWRLRHNIGASYSISLT 447

RESULT 9
O979N3 ID O979N3 PRELIMINARY: PRT: 116 AA.
AC O979N3:
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein TV1127.
GN TV1127 OR TVG156690.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmatata; Thermoplasmatales;

```

```

OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed-11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Ninoshita T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL: AP000995; BAB60269.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 116 AA; 13804 MW; 895956AD82F96801 CRC64;

Query Match
Best Local Similarity 32.7%; Score 50; DB 17; Length 116;
Matches 11; Conservative 6; Mismatches 5; Indels 4; Gaps 1;

QY 5 KLRTHTGTTGQD---YFKFTLLAD 26
DB 32 RLDPYRSTGDRYDYFNFRTLLD 57

RESULT 10
082SA4 PRELIMINARY; PRT; 161 AA.
AC 082SA4;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein A117607.
GN AL17607.
OS Anabaena sp. (strain PCC 7120).
OC Plasmid PCC7120beta.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed-11759840;
RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003602; BAB77250.1; -.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 161 AA; 17490 MW; 295C5A82F64DE453 CRC64;

Query Match
Best Local Similarity 32.7%; Score 50; DB 16; Length 161;
Matches 9; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 2 EREKLRVHTGDTGOYEFRTLLADDS 29
DB 122 EENNMYKTAVTGDKFYKSTILLGMNA 149

RESULT 11
095QJ2 PRELIMINARY; PRT; 521 AA.
AC 095QJ2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical 57.7 kDa protein.
GN F25B5.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

```

```

OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Tach A.;
RT "The sequence of C. elegans cosmid F25B5."
RL Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: U23172; AL024494.1; -.
DR InterPro: IPR001645; Fpolylgl-synthase.
DR InterPro: IPR000713; Mur_1lgase.
DR InterPro: IPR004101; Mur_1lgase.C.
DR Pfam: PF01225; Mur_1lgase.1.
DR PROSITE: PS01011; FOLYLPOLYGLU_SYNTH_1; UNKNOWN.1.
DR PROSITE: PS01012; FOLYLPOLYGLU_SYNTH_2; UNKNOWN.1.
KW Hypothetical protein.
SQ SEQUENCE 521 AA; 57655 MW; CC15C5CF6FCA351D CRC64;

Query Match
Best Local Similarity 31.7%; Score 48.5; DB 5; Length 521;
Matches 13; Conservative 5; Mismatches 5; Indels 23; Gaps 1;

QY 3 REKLRVHTGTTGQD-----YFKFTLLA 25
DB 140 REKIVDGGPVSEQFAEFPHVYDIKREHSDMNPVFFRTLLA 185

RESULT 12
061080 PRELIMINARY; PRT; 1186 AA.
AC 061080;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Myosin IC heavy chain.
GN MICHC.
OS Acanthamoeba castellanii (Acanthamoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88016163; PubMed-3477803;
RA Jung G., Korn E.D., Hammer J.A. III;
RT "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
RT and non-myosin-like sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99079990;
RA Wang Z.Y., Wang F., Sellers J.R., Korn E.D., Hammer J.A. III;
RT "Analysis of the regulatory phosphorylation site in Acanthamoeba
RT myosin IC by using site-directed mutagenesis."
RL Proc. Natl. Acad. Sci. U.S.A. 95:15200-15205(1998).
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AF051353; AAC98089.1; -.
DR HSSP: P08799; 1MND.
DR InterPro: IPR001609; myosin_head.
DR InterPro: IPR001452; SH3.

```

```

DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000066; SH3; 1.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00242; MYSC; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain.
SQ
SEQUENCE 1186 AA; 129459 MW; E37AD44A685803A6 CRC64;

Query Match
Best Local Similarity 31.7%; Score 48.5; DB 5; Length 1186;
Matches 13; Conservative 3; Mismatches 9; Indels 1; Gaps 1;

QY 1 YEREKLRVTHQTGDQYFKFTLLADQHS 25
DB 181 YLEKSRVYVYQTNGERNHFIFQLLA 206

RESULT 13
Q805B5 PRELIMINARY; PRT; 50 AA.
AC Q805B5;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE AGR_C_2616p.
GN AGR_C_2616p.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OC NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Askenazi M., Halling C., Mullin L.,
RA Ouriello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA Wollem C., Allinger M., Doughy D., Scott C., Lappas C., Markez B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AE008067; AAK87209.1; -.
SQ SEQUENCE 50 AA; 5660 MW; AAF6E9AC6927A9F CRC64;

Query Match
Best Local Similarity 31.4%; Score 48; DB 16; Length 50;
Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 7 RVTHQTGDQYFKFTLLADQHS 29
DB 7 RFSHATSGDQSFLLIVTTSQTS 29

RESULT 14
Q29564 PRELIMINARY; PRT; 190 AA.
AC Q29564;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE Replication control protein A, putative.
GN AF0694.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
DR EMBL: AE001057; AAB90548.1; -.
DR TIGR: AF0694; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 190 AA; 21896 MW; B16AD07E23F52A73 CRC64;

Query Match
Best Local Similarity 31.4%; Score 48; DB 17; Length 190;
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 REKLRVTHQTGDQYFKFTLLADQHS 28
DB 128 RDRIVVYFTTNDRIFDFLTLLADRY 153

RESULT 15
Q8YM49 PRELIMINARY; PRT; 318 AA.
AC Q8YM49;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein A115091.
GN A115091.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OC NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneo T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003596; BAB76790.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 318 AA; 34767 MW; B8BF559AE2DB681 CRC64;

Query Match
Best Local Similarity 31.4%; Score 48; DB 16; Length 318;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 YEREKLRVTHQTGDQYFKF 20
DB 94 FERKQAQLTSQVTPHVLNF 113

Search completed: January 29, 2003, 06:50:00
Job time : 67.1053 secs

```

3

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 06:44:44 ; Search time 30.4561 Seconds
(without alignments)
122.505 Million cell updates/sec

Title: US-09-627-165D-20

Perfect score: 145

Sequence: 1 YEREKLVHTQTGTDEYFRFTLLADTV 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: A_Geneseq.101002:*
2: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1980.DAT.*
3: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1981.DAT.*
4: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1982.DAT.*
5: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1983.DAT.*
6: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1984.DAT.*
7: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1985.DAT.*
8: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1986.DAT.*
9: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1987.DAT.*
10: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1988.DAT.*
11: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1989.DAT.*
12: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1990.DAT.*
13: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1991.DAT.*
14: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1992.DAT.*
15: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1993.DAT.*
16: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1994.DAT.*
17: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1995.DAT.*
18: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1996.DAT.*
19: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1997.DAT.*
20: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1998.DAT.*
21: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA1999.DAT.*
22: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA2000.DAT.*
23: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA2001.DAT.*
24: /SID22/gcgcdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	122	84.1	28	15	AA14926
2	122	84.1	254	22	AA147090
3	122	84.1	256	20	AA125981
4	122	84.1	256	20	AA125984
5	122	84.1	551	23	AA179450
6	122	84.1	564	18	AA110021
7	122	84.1	564	20	AA190127
8	119	82.1	256	22	AA147092
9	118	81.4	252	19	AA146559
10	118	81.4	252	19	AA146561

11	118	81.4	253	18	AA110022
12	118	81.4	253	20	AA190125
13	118	81.4	254	20	AA125980
14	118	81.4	254	20	AA125983
15	118	81.4	531	20	AA125979
16	118	81.4	532	20	AA125982
17	116	80.0	254	22	AA147091
18	115	79.3	255	20	AA125971
19	115	79.3	255	20	AA125977
20	115	79.3	255	20	AA125970
21	115	79.3	533	20	AA125973
22	115	79.3	533	20	AA125976
23	115	79.3	533	20	AA125979
24	94	64.8	20	15	AA14925
25	50	34.5	1120	23	AB107425
26	49.5	34.1	36	23	AB102301
27	49	33.8	9	20	AA190128
28	48	33.1	758	22	AB183337
29	48	33.1	1560	21	AA167203
30	48	33.1	1561	21	AA177202
31	48	33.1	1562	21	AA186339
32	48	33.1	1562	21	AA177194
33	48	33.1	12199	21	AA177180
34	47.5	32.8	654	21	AA130309
35	47.5	32.8	673	21	AA130308
36	47.5	32.8	688	21	AA130307
37	46	31.7	253	20	AA137029
38	46	31.7	513	22	AB124916
39	45.5	31.4	1326	22	AB159965
40	45	31.0	1042	18	AA119765
41	45	31.0	1042	18	AA119765
42	44.5	30.7	176	22	AB165319
43	44	30.3	74	23	AB109030
44	44	30.3	125	23	AB143875
45	44	30.3	141	23	AB142622

ALIGNMENTS

RESULT 1
AAR44926 standard; peptide; 28 AA.
AAR44926:
04-JUL-1994 (first entry)
Galactoside-specific lectin ML-I A2 chain N-terminal.
Mistletoe; immunomodulatory; cancer; therapy; treatment; adjuvant;
supportive; palliative; toxic.
Viscum album.
DEF4221836-A.
05-JAN-1994.
03-JUL-1992; 92DE-4221836.
PR 03-JUL-1992; 92DE-4221836.
PA (BARD/) BARDOSI A.
PA (GABI/) GABIUS HJ.
WPI: 1994-008729/02.
New mistletoe lectin - with immunomodulatory activity, useful for
cancer adjuvant therapy
Claim 3; Page 5; Bpp: German.

Prepro mistletoe I
Mistletoe ML-A-cha
Mistletoe lectin A
Mistletoe lectin A
Mistletoe lectin A
Mistletoe lectin I
Mistletoe lectin I
A-chain isoform fo
Mistletoe lectin A
Mistletoe lectin A
Mistletoe lectin A
Mistletoe lectin p
Mistletoe lectin p
Mistletoe lectin p
Galactoside-specific
Novel human diagno
Human ORF protein
Mistletoe lectin M
Drosophila melanog
Narbonolide synthase
S. venezuelae plk
Amino acid sequenc
S. venezuelae macr
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Protein involved I
Novel human diagno
Drosophila melanog
Interleukin-
Human SRA-1 protel
Drosophila melanog
Human ORF protein
Human peptide enco
Human ovarian anti

KM lymphokine-producing macrophage; uncontrolled cell growth; treatment;
 KM cancer; cytotoxicity; antigen; isoform; lectin A2.
 OS Viscum album.
 PN DE19804210-A1.
 XX 12-AUG-1999.
 PD 03-FEB-1998; 98DE-1004210.
 XX 03-FEB-1998; 98DE-1004210.
 PR 03-FEB-1998; 98DE-1004210.
 PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
 PI Morris P, Stiefel T, Voelter W, Welters P;
 DR WPI: 1999-445335/38.
 DR N-PSDB: AA209108.
 XX
 XX Preparation of mistletoe lectins in heterologous systems,
 PT particularly for use as anticancer agents and immunostimulants
 PS
 XX Disclosure: Fig 6B; 78pp; German.
 CC This invention describes a novel mistletoe lectin (I) and its fragments
 CC which have antitumour and immunostimulatory activity. The A-chain (MLA)
 CC of the mistletoe lectin binds to, and inactivates, the 28S subunit of
 CC ribosomes. Non-cytotoxic forms of (I) activate T-cell and
 CC lymphokine-producing macrophages, so stimulate immunity. (I) and its
 CC fragments are used to treat uncontrolled cell growth (particularly
 CC cancers) and if they lack cytotoxicity, to increase the strength of the
 CC immune response, particularly to a co-administered antigen
 CC (tumour-associated, bacterial or viral). The method allows production of
 CC mistletoe lectin, and its individual chains, in many different isoforms
 CC and on a large scale, at any time of the year. Recombinant products are
 CC free from toxins present in natural mistletoe extracts. This sequence
 CC represents a fragment of a mistletoe lectin A2 protein variant.
 SQ Sequence 256 AA:
 QY 1 YEREKLRVTHQTGDEYFRFTLLRDYV 28
 DB 1 YERLRLRVTHQTGDEYFRFTLLRDYV 28
 RESULT 5
 ABB79450
 ID ABB79450 standard; Protein; 551 AA.
 AC ABB79450;
 XX
 DT 08-JUL-2002 (first entry)
 XX
 DE Galactose-recognising mistletoe lectin.
 XX
 KM Mistletoe; galactose-recognising mistletoe lectin; MLIII.
 XX
 OS Viscum album.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 223 /note- "Encoded by ATG"
 FT Misc-difference 251 /note- "Encoded by TTT"
 FT Misc-difference 344 /note- "Encoded by TCG"
 FT Misc-difference 380 /note- "Encoded by GCC"
 FT

FT Misc-difference 448 /note- "Encoded by GTC"
 FT
 XX
 XX DE10044027-A1.
 PN 14-MAR-2002.
 XX
 PD 06-SEP-2000; 2000DE-1044027.
 XX
 PF 06-SEP-2000; 2000DE-1044027.
 PR 06-SEP-2000; 2000DE-1044027.
 XX
 PA (VISC-) VISCUM AG.
 PI Kieff S;
 DR WPI: 2002-316737/36.
 DR N-PSDB: ABL56947.
 XX
 XX New nucleic acid encoding preprotein of mistletoe lectin, useful as
 PT diagnostic and therapeutic agents, also encodes polypeptide -
 PS
 XX Claim 1; Fig 1; 6pp; German.
 CC The invention relates to a nucleic acid molecule (ABL56947) that encodes
 CC a preprotein (ABB79450) which, after maturation, has the biological
 CC activity of the galactose-recognising mistletoe lectin (MLIII). The MLIII
 CC encoding nucleic acid molecule, primers specific to it or complements of
 CC it, and encoded (oligomeric) polypeptides are useful as diagnostic and
 CC therapeutic agents.
 SQ Sequence 551 AA:
 QY 1 YEREKLRVTHQTGDEYFRFTLLRDYV 28
 DB 34 YERLRLRVTHQTGDEYFRFTLLRDYV 61
 RESULT 6
 AAM10021
 ID AAM10021 standard; Protein; 564 AA.
 AC AAM10021;
 XX
 DT 18-DEC-1997 (first entry)
 XX
 DE Prepro mistletoe lectin.
 XX
 KM Mistletoe; lectin; cytotoxic; A chain; B chain; dimer.
 XX
 OS Viscum album.
 XX
 PN EP751221-A1.
 PD 02-JAN-1997.
 XX
 PF 26-JUN-1995; 95EP-0109949.
 PR 26-JUN-1995; 95EP-0109949.
 XX
 PA (MADU) MADUS KOELN AG.
 PI Baur A, Eck J, Lentzen H, Zinke H;
 DR WPI: 1997-054678/06.
 DR N-PSDB: AAT70473.
 XX
 PT Nucleic acid encoding pre:pro form of mistletoe lectin - for
 XX therapeutic or diagnostic use

```

PS Claim 12; Fig 4C; 30pp; German.
XX
CC Mistletoe lectin is a cytotoxic agent that has been used for tumour
CC therapy. It can be used in immunotoxins and medicaments. Nucleic
CC acid fragments can be used in diagnostic methods. Mistletoe lectin (
CC AAT70473) comprises an A chain (AAT70474) and a B chain (AAT70475).
XX
SQ Sequence 564 AA;

Query Match      84.1%; Score 122; DB 18; Length 564;
Best Local Similarity 85.7%; Pred. No. 1.6e-10;
Matches 24; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YEREKLRVTHQTGDEYFRFTLLADTV 28
   ||| :|||||
Db 34 YERLRRLRVTHQTGDEYFRFTLLRDYV 61

RESULT 7
ID AAM90127 standard; Protein; 564 AA.
XX
AC AAM90127;
XX
DT 30-APR-1999 (first entry)
XX
DE Mistletoe lectin prepro-protein.
XX
KM ML; mistletoe; lectin; ML; transgenic plant; glycosylation;
KW dimer; immunotoxin; large-scale production; diagnosis; therapeutic;
KM cancer.
XX
OS Viscum album.
XX
PN EP884388-A1.
XX
PD 16-DEC-1998.
XX
PF 26-JUN-1995; 95EP-0109949.
XX
PR 26-JUN-1995; 95EP-0109949.
XX
PR 26-JUN-1995; 98EP-0105660.
XX
PA (MADU ) MADAUUS KOELN AG.
XX
PI Baur A, Eck J, Lentzen H, Zinke H;
XX
DR WPI: 1999-026582/03.
DR N-PSDB: AAV74182.
XX
PT New transgenic plant expressing mistletoe lectin - useful for
PT producing recombinant lectin in e.g. cancer diagnosis and therapy
XX
PS Claim 1a; Fig 4c; 30pp; German.
XX
CC This invention describes a novel transgenic plant transformed with a
CC vector capable of encoding a mistletoe (Viscum album) lectin
CC preproprotein or a biologically active fragment. The specification
CC also describes a polypeptide produced by a plant where the polypeptide
CC exhibits at least one enzymatic modification other than the glycosylation
CC that occurs in viscum album or the polypeptide is a fusion protein, a
CC mistletoe lectin polypeptide dimer and an immunotoxin comprising the
CC polypeptide or the polypeptide dimer. The plants are used for large-scale
CC production of mistletoe lectin for diagnostic or therapeutic purposes
CC (e.g. in cancer therapy). This sequence represents the mistletoe lectin
CC used in the method of the invention.
XX
SQ Sequence 564 AA;

Query Match      84.1%; Score 122; DB 20; Length 564;
Best Local Similarity 85.7%; Pred. No. 1.6e-10;
Matches 24; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 YEREKLRVTHQTGDEYFRFTLLADTV 28
   ||| :|||||
Db 34 YERLRRLRVTHQTGDEYFRFTLLRDYV 61

RESULT 8
ID AAB47092 standard; Protein; 256 AA.
XX
AC AAB47092;
XX
DT 16-MAY-2001 (first entry)
XX
DE A-chain isoform for biosynthesis of a Korean mistletoe lectin #3.
XX
KM Isoform: A-chain; B-chain; biosynthesis; lectin; Korean mistletoe;
KW KML; tumour; KM-110; KML-C; KWHBP; KML-11U; KML-11L;
KW heparin binding protein.
XX
OS Viscum album coloratum.
XX
PN EP1074560-A2.
XX
PD 07-FEB-2001.
XX
PF 27-JUL-2000; 2000EP-0402168.
XX
PR 27-JUL-1999; 99KR-0030638.
XX
PR (MIST-) MISTLE BIOTECH CO LTD.
XX
PI Kim J, Song S, Suh B, Lee K, Doo M, Kwak J, Song B, Yoon T;
PI Kang T, Park C;
XX
DR WPI: 2001-171044/18.
DR N-PSDB: AAC85474.
XX
PT Novel lectin proteins isolated from Korean mistletoe, useful for
PT enhancing immunity and effectuating anti-tumoral activity -
XX
PS Claim 6; Page 28; 62pp; English.
XX
CC The sequences given in AAB47090-92 are isoforms of an A-chain gene
CC involved in biosynthesis of lectins isolated from Korean mistletoe.
CC Korean mistletoe lectins (KML) are useful for enhancing immunity and
CC for treating tumours. The KML's are isolated from a protein fraction
CC derived from the leaves, stems and fruits of Korean mistletoe, which
CC is designated KM-110. One of the isolates, KML-C was shown to be
CC effective against colon 26-M3.1 carcinoma and LS178Y-MU25 lymphoma.
XX
SQ Sequence 256 AA;

Query Match      82.1%; Score 119; DB 22; Length 256;
Best Local Similarity 82.1%; Pred. No. 1.9e-10;
Matches 23; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YEREKLRVTHQTGDEYFRFTLLADTV 28
   ||| :|||||
Db 1 YERLRRLRVTHQTGDEYFRFTLLRDYV 28

RESULT 9
ID AAM64659 standard; Protein; 252 AA.
XX
AC AAM64659;
XX
DT 23-OCT-1998 (first entry)
XX
DE Mistletoe rMLA protein.
XX
KW Lectin A-chain; mistletoe; rMLA; fusion protein; effector; cytotoxic;
KW intracellular; processing module; protease recognition; targeting module;

```

KW Internalisation; treatment; disorder; cell proliferation; activation;
 KW autoimmune disease; allergy; tumour; ricin; translocation.
 OS Viscum album.
 XX W09829540-A2.
 XX
 XX 09-JUL-1998.
 XX
 XX 02-JAN-1998; 98MO-EP00009.
 XX
 XX 02-JAN-1997; 97EP-0100012.
 XX
 XX (BRAI-) BRAIN BIOTECHNOLOGY RES & INFORMATION NE.
 XX
 XX Eck J, Schmidt A, Zinke H;
 XX WPI: 1998-388122/33.
 XX N-PSDB; AAV51341.
 XX
 XX Nucleic acid encoding fusion protein containing mistletoe lectin A
 PT chain - useful for treatment of proliferative and autoimmune
 PT diseases, allergies and tumours
 XX
 XX Disclosure: Fig 11a; 115pp; German.
 PS
 XX This sequence represents a lectin A-chain, rMLA, isolated from mistletoe.
 CC This sequence can be used in the construction of a fusion protein which
 CC comprises an effector module that is cytotoxic intracellularly, a
 CC processing module covalently bonded to the effector module and
 CC containing a protease recognition sequence, and a targeting module
 CC covalently bonded to the processing module, able to bind specifically to
 CC the surface of a cell so as to mediate internalisation of the fusion
 CC protein. Such a fusion protein can be used for treating disorders
 CC involving proliferation and/or elevated activation of cells, especially
 CC autoimmune disease, allergy and tumours. The proteins can be administered
 CC e.g. by injection or topically but especially by intravenous injection,
 CC at 1 ng to 500 mu g/kg/day, or for ex vivo use at 1 pg to 500 ng/ml.
 CC Fusion proteins can develop toxic activity in a wide range of target
 CC cells. The processing module prevents extracellular dissociation, and
 CC fusion proteins based on mistletoe lectin A-chain are far more active
 CC than those based on ricin and do have the associated problems of
 CC non-specific toxicity. The protein may be expressed in a non-glycosylated
 CC form that does not bind to sugar receptors in the liver, and which has a
 CC long half-life in the blood. Where the mistletoe lectin B-chain is used,
 CC it actively assists in translocation of the ML A-chain from the
 CC endoplasmic reticulum to the cytoplasm.
 CC
 CC Sequence 252 AA:
 SQ
 Query Match 81.4%; Score 118; DB 19; Length 252;
 Best Local Similarity 82.1%; Pred. No. 2.6e-10;
 Matches 23; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 YEREKLRVTHQTGDEYFRITLLADTV 28
 |||:|||||||:|||||||:|
 Db 2 YERIRLRVTHQTGDEYFRITLLRDYV 29
 RESULT 10
 AAW64661
 ID AAW64661 standard; Protein; 252 AA.
 XX
 AC AAW64661;
 XX
 DT 23-OCT-1998 (first entry)
 XX
 DE Mistletoe rMLA variant protein.
 XX
 KW Lectin A-chain; mistletoe; rMLA; fusion protein; effector; cytotoxic;
 KW intracellular; processing module; protease recognition; targeting module;
 KW internalisation; treatment; disorder; cell proliferation; activation;
 KW autoimmune disease; allergy; tumour; ricin; translocation; ss.

XX
 OS Viscum album.
 XX
 XX Key Location/Qualifiers
 FT Protein 1..252
 FT /note="partial"
 XX
 XX W09829540-A2.
 XX
 XX 09-JUL-1998.
 XX
 XX 02-JAN-1998; 98MO-EP00009.
 XX
 XX 02-JAN-1997; 97EP-0100012.
 XX
 XX (BRAI-) BRAIN BIOTECHNOLOGY RES & INFORMATION NE.
 XX
 XX Eck J, Schmidt A, Zinke H;
 XX WPI: 1998-388122/33.
 XX N-PSDB; AAV51343.
 XX
 XX Nucleic acid encoding fusion protein containing mistletoe lectin A
 PT chain - useful for treatment of proliferative and autoimmune
 PT diseases, allergies and tumours
 XX
 XX Disclosure: Fig 11a; 115pp; German.
 PS
 XX This sequence encodes a variant mistletoe lectin A-chain, rMLA. This
 CC sequence can be used in the construction of a fusion protein which
 CC comprises an effector module that is cytotoxic intracellularly, a
 CC processing module covalently bonded to the effector module and
 CC containing a protease recognition sequence, and a targeting module
 CC covalently bonded to the processing module, able to bind specifically to
 CC the surface of a cell so as to mediate internalisation of the fusion
 CC protein. Such a fusion protein can be used for treating disorders
 CC involving proliferation and/or elevated activation of cells, especially
 CC autoimmune disease, allergy and tumours. The proteins can be administered
 CC e.g. by injection or topically but especially by intravenous injection,
 CC at 1 ng to 500 mu g/kg/day, or for ex vivo use at 1 pg to 500 ng/ml.
 CC Fusion proteins can develop toxic activity in a wide range of target
 CC cells. The processing module prevents extracellular dissociation, and
 CC fusion proteins based on mistletoe lectin A-chain are far more active
 CC than those based on ricin and do have the associated problems of
 CC non-specific toxicity. The protein may be expressed in a non-glycosylated
 CC form that does not bind to sugar receptors in the liver, and which has a
 CC long half-life in the blood. Where the mistletoe lectin B-chain is used,
 CC it actively assists in translocation of the ML A-chain from the
 CC endoplasmic reticulum to the cytoplasm.
 CC
 CC Sequence 252 AA:
 SQ
 Query Match 81.4%; Score 118; DB 19; Length 252;
 Best Local Similarity 82.1%; Pred. No. 2.6e-10;
 Matches 23; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 YEREKLRVTHQTGDEYFRITLLADTV 28
 |||:|||||||:|||||||:|
 Db 1 YERIRLRVTHQTGDEYFRITLLRDYV 28
 RESULT 11
 AAW10022
 ID AAW10022 standard; Protein; 253 AA.
 XX
 AC AAW10022;
 XX
 DT 18-DEC-1997 (first entry)
 XX
 DE Prepro mistletoe lectin A chain.
 XX
 KW Mistletoe; lectin; cytotoxic; A chain; B chain; dimer.
 XX

OS Viscum album.
 XX
 XX EP751221-A1.
 PN
 XX 02-JAN-1997.
 PD
 XX
 XX 26-JUN-1995; 95EP-0109949.
 PF
 XX 26-JUN-1995; 95EP-0109949.
 PR
 XX 26-JUN-1995; 95EP-0109949.
 XX
 PA (MADU) MADDAUS KOELN AG.
 PI Baur A, Eck J, Lentzen H, Zinke H;
 XX
 XX WPI: 1997-054678/06.
 DR N-PSDB; AAT70474.
 DR
 XX Nucleic acid encoding pre:pro form of mistletoe lectin - for
 PT therapeutic or diagnostic use
 PS
 XX Claim 12; Fig 4A; 30pp; German.
 CC Mistletoe lectin is a cytotoxic agent that has been used for tumour
 CC therapy. It can be used in immunotoxins and medicaments. Nucleic
 CC acid fragments can be used in diagnostic methods. Mistletoe lectin (
 CC AAT70473) comprises an A chain (AAT70474) and a B chain (AAT70475).
 CC
 XX
 SQ Sequence 253 AA;
 Query Match 81.4%; Score 118; DB 18; Length 253;
 Best Local Similarity 82.1%; Pred. No. 2.7e-10;
 Matches 23; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 YEREKLRVTHQTGDEYFRFTLLADTV 28
 ||| :|||||||:||||||| | |
 Db 2 YERIRLRVTHQTGEYFRFTLLRDYV 29
 RESULT 12
 AAW90125
 ID AAW90125 standard; Protein; 253 AA.
 XX
 AC AAW90125;
 XX
 DT 30-APR-1999 (first entry)
 DE
 XX Mistletoe ML A-chain protein.
 KW ML; mistletoe; lectin; MIA; A-chain; transgenic plant; glycosylation;
 KW dimer; immunotoxin; large-scale production; diagnosis; therapeutic;
 KW cancer.
 XX
 XX Viscum album.
 OS
 XX
 PN EP864388-A1.
 PD 16-DEC-1998.
 PD
 XX 26-JUN-1995; 95EP-0109949.
 PF
 XX 26-JUN-1995; 95EP-0109949.
 PR
 XX 26-JUN-1995; 98EP-0105660.
 XX
 PA (MADU) MADDAUS KOELN AG.
 PI Baur A, Eck J, Lentzen H, Zinke H;
 XX
 XX WPI: 1999-026582/03.
 DR N-PSDB; AAV74180.
 DR
 XX New transgenic plant expressing mistletoe lectin - useful for
 PT producing recombinant lectin in e.g. cancer diagnosis and therapy
 XX

PS Disclosure; Fig 4a; 30pp; German.
 XX
 CC This invention describes a novel transgenic plant transformed with a
 CC vector capable of encoding a mistletoe (Viscum album) lectin
 CC preproprotein or a biologically active fragment. The specification
 CC also describes a polypeptide produced by a plant where the polypeptide
 CC exhibits at least one enzymatic modification other than the glycosylation
 CC that occurs in Viscum album or the polypeptide is a fusion protein, a
 CC mistletoe lectin polypeptide dimer and an immunotoxin comprising the
 CC polypeptide or the polypeptide dimer. The plants are used for large-scale
 CC production of mistletoe lectin for diagnostic or therapeutic purposes
 CC (e.g. in cancer therapy). This sequence represents the mistletoe lectin
 CC A-chain which is contained in expression vector pT7MA.
 CC
 XX
 SQ Sequence 253 AA;
 Query Match 81.4%; Score 118; DB 20; Length 253;
 Best Local Similarity 82.1%; Pred. No. 2.7e-10;
 Matches 23; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 YEREKLRVTHQTGDEYFRFTLLADTV 28
 ||| :|||||||:||||||| | |
 Db 2 YERIRLRVTHQTGEYFRFTLLRDYV 29
 RESULT 13
 AAT25980
 ID AAT25980 standard; Protein; 254 AA.
 XX
 AC AAT25980;
 XX
 DT 18-OCT-1999 (first entry)
 DE
 XX Mistletoe lectin A1 protein fragment.
 KW Mistletoe; lectin; antitumour; immunostimulant; A-chain; MIA; immunity;
 KW ribozyme 28S subunit; non-cytotoxic; T-cell activation; immune response;
 KW lymphokine-producing macrophage; uncontrolled cell growth; treatment;
 KW cancer; cytotoxicity; antigen; isoform; lectin A1.
 XX
 OS Viscum album.
 XX
 PN DE19804210-A1.
 PD 12-AUG-1999.
 PD
 XX 03-FEB-1998; 98DE-1004210.
 PF
 XX 03-FEB-1998; 98DE-1004210.
 PR
 XX 03-FEB-1998; 98DE-1004210.
 XX
 PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
 PI Morris P, Stiefel T, Voelter W, Welters P;
 XX
 XX WPI: 1999-445335/38.
 DR N-PSDB; AA209104.
 DR
 XX Preparation of mistletoe lectins in heterologous systems,
 PT particularly for use as anticancer agents and immunostimulants
 PS
 XX Disclosure; Fig 2b; 78pp; German.
 CC
 CC This invention describes a novel mistletoe lectin (I) and its fragments
 CC which have antitumour and immunostimulatory activity. The A-chain (MIA)
 CC of the mistletoe lectin binds to, and inactivates, the 28S subunit of
 CC ribosomes. Non-cytotoxic forms of (I) activate T-cell and
 CC lymphokine-producing macrophages, so stimulate immunity. (II) and its
 CC fragments are used to treat uncontrolled cell growth (particularly
 CC cancers) and if they lack cytotoxicity, to increase the strength of the
 CC immune response, particularly to a co-administered antigen
 CC (tumour-associated, bacterial or viral). The method allows production of
 CC mistletoe lectin, and its individual chains, in many different isoforms
 CC and on a large scale, at any time of the year. Recombinant products are

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 06:44:49 : Search time 11.2982 Seconds
(without alignments)
72.918 Million cell updates/sec

Title: US-09-627-165D-20

Perfect score: 145

Sequence: 1 YEREKLRVTHQTGDEYFRFTLLADTV 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.AA.*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122	84.1	564	4	US-08-776-059-35
2	118	81.4	253	4	US-08-776-059-31
3	81	55.9	50	4	US-08-776-059-37
4	77	53.1	18	4	US-08-776-059-52
5	48	33.1	1562	3	US-09-320-878-3
6	48	33.1	1562	3	US-09-105-537-35
7	48	33.1	11877	4	US-09-105-537-6
8	45	31.0	747	4	US-09-362-336A-14
9	45	31.0	903	2	US-08-209-521-24
10	45	31.0	903	4	US-08-961-810-134
11	45	31.0	903	4	US-08-352-902D-134
12	44	30.3	129	3	US-08-513-974B-327
13	44	30.3	352	3	US-08-513-974B-374
14	44	30.3	352	3	US-08-513-974B-374
15	43	29.7	545	4	US-08-687-590-26
16	43	29.7	240	2	US-08-380-403A-6
17	43	29.7	240	2	US-08-895-628-6
18	43	29.7	510	5	US-08-895-810D-6
19	43	29.7	633	5	PCT-US96-03916-3
20	43	29.7	633	2	US-08-380-403A-2
21	43	29.7	633	2	US-08-380-403A-5
22	43	29.7	633	2	US-08-895-628-2
23	43	29.7	633	2	US-08-895-628-5
24	43	29.7	633	4	US-08-895-810D-2
25	43	29.7	633	4	US-08-895-810D-5
26	43	29.7	3724	2	US-08-804-227C-10
27	43	29.7	3724	2	US-08-804-198-4

28	42	29.0	269	4	US-08-713-556F-38	Sequence 38, Appl
29	42	29.0	373	2	US-08-559-524A-4	Sequence 4, Appl1
30	42	29.0	373	2	US-08-749-707-4	Sequence 4, Appl1
31	42	29.0	437	5	PCT-US96-10043-11	Sequence 11, Appl
32	42	29.0	1157	1	US-07-876-280-30	Sequence 30, Appl
33	42	29.0	1157	1	US-07-812-180A-2	Sequence 2, Appl1
34	42	29.0	1157	1	US-08-315-468-2	Sequence 2, Appl1
35	42	29.0	1157	1	US-07-941-650A-2	Sequence 2, Appl1
36	41	28.3	9	4	US-08-776-059-46	Sequence 46, Appl
37	41	28.3	64	4	US-08-936-165A-447	Sequence 47, Appl
38	41	28.3	94	3	US-09-147-550-16	Sequence 16, Appl
39	41	28.3	94	3	US-09-147-550-24	Sequence 24, Appl
40	41	28.3	94	3	US-09-147-550-34	Sequence 34, Appl
41	41	28.3	94	3	US-09-147-550-38	Sequence 38, Appl
42	41	28.3	94	3	US-09-147-550-52	Sequence 52, Appl
43	41	28.3	94	3	US-09-147-550-57	Sequence 57, Appl
44	41	28.3	94	3	US-09-147-550-61	Sequence 61, Appl
45	41	28.3	94	3	US-09-147-550-71	Sequence 71, Appl

ALIGNMENTS

RESULT 1

US-08-776-059-35
Sequence 35, Application US/08776059B
Patent No. 6271368

GENERAL INFORMATION:
APPLICANT: LENTZEN, Hans

Query Match 84.1% Score 122. DB 4; Length 564;
Best Local Similarity 85.7% Pred. No. 7.6e-12;
Matches 24; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 YEREKLRVTHQTGDEYFRFTLLADTV 28
DB 34 YERLRLRVTHQTGDEYFRFTLLADTV 61

RESULT 2
US-08-776-059-31
Sequence 31, Application US/08776059B
Patent No. 6271368

GENERAL INFORMATION:
APPLICANT: LENTZEN, Hans

```

: EARLIER FILING DATE: 1995-06-26
: NUMBER OF SEQ ID NOS: 56
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 31
: LENGTH: 253
: TYPE: PRT
: ORGANISM: Viscum album
US-08-776-059-31

```

Query Match	81.4%;	Score 118;	DB 4;	length 253;
Best Local Similarity	82.1%;	Pred. No. 1.4e-11;		
Matches 23; Conservative		2; Mismatches 3;	Indels 0;	Gaps 0;

```
QY      1 YEREKLRVTHQOTGDEYERFITLLADTV 28
        ||| : ||||| ||| : ||||| ||| | |
Db      2 YERIRLRVTHQOTGGEYERFITLLRDYV 29
```

```

US-08-776-059-37
RESULT 3
; Sequence 37, Application US/08776059B
; Patent No. 6271368
; GENERAL INFORMATION:
; APPLICANT: LENTZEN, Hans
; APPLICANT: ECK, Jürgen
; APPLICANT: BAUER, Axel
; APPLICANT: ZINKE, Holger
; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)
; FILE REFERENCE: 674503-2003
; CURRENT APPLICATION NUMBER: US/08/776, 059B
; EARLIER FILING DATE: 1999-06-19
; EARLIER APPLICATION NUMBER: PCT/EP96/02273
; EARLIER FILING DATE: 1996-06-25
; EARLIER APPLICATION NUMBER: 95109949.8
; EARLIER FILING DATE: 1995-06-26
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 37
; LENGTH: 50
; TYPE: prt
; ORGANISM: Viscum album
US-08-776-059-37
```

Query Match	55.98;	Score 81;	DB 4;	Length 50;
Best Local Similarity	88.28;	Pred. No. 2.3e-06;		
Matches 15; Conservative	1;	Mismatches 1;	Indels 0;	Gaps 0;

```

QY      1 VERKLRVTHOTGDEY 17
      ||| :|||||||
DB      34 VERKLRVTHOTGDEY 50

RESULT 4
US-08-776-059-52
; Sequence 52, Application US/08776059B
; Patent No. 6271368
;
; GENERAL INFORMATION:
;
; APPLICANT: LENTZEN, Hans
;
; APPLICANT: ECK, Jürgen
;
; APPLICANT: BAUR, Axel
;
; APPLICANT: ZINKE, Holger
;
; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)
;
; FILE REFERENCE: 674503-2003
;
; CURRENT APPLICATION NUMBER: US/08/776,059B
;
; CURRENT FILING DATE: 1999-06-19
;
; EARLIER APPLICATION NUMBER: PCT/EP96/02273
;
; EARLIER FILING DATE: 1996-06-25
;
; EARLIER APPLICATION NUMBER: 95109949.8
;
; EARLIER FILING DATE: 1995-06-26
;
; NUMBER OF SEQ ID NOS: 56
;
; SOFTWARE: Patentlin Ver. 2.0
;
; SEQ ID NO 52
;
; LENGTH: 18

```

```

; TYPE: PRT
; ORGANISM: Spodoptera frugiperda
US-08-776-059-52

```

```
Oy      1 YEREKLRVTHQTGDEY 17
        ||| : ||||| ||| : ||
Db      2 YERIRLRVTHQTGGEY 18
```

```

RESULT 5
US-09-320-878-3
: Sequence 3, Application US/09320878A
: Patent No. 6117659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 300622002120
: CURRENT APPLICATION NUMBER: US/09/320,878A
: CURRENT FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141,908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073,538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846,247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119,139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100,880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087,080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 1562
: TYPE: prt
: ORGANISM: Streptomyces venezuelae
US-09-320-878-3

```

Query Match	Similarity	33.1%	Score	48	DB	3	Length	1562	
Best Local	Similarity	44.8%	Pred. No.	30					
Matches	13	Conservative	5	Mismatches	9	Indels	2	Gaps	1
Oy	2	EREKL--RVYHOTGDEYFFFIITLADTV	28						
		: : : : : : : : : : : : : : : :							
Db	970	DREALAARLTATLTGDEGTGVSLDDLV	998						

```

RESULT 6--35
US-09-105-537-35
; Sequence 35 Application US/0910537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 1562

```


STATE: Oregon
COUNTRY: U.S.A.
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,810
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Van Rysselberghe, Pierre C.
REGISTRATION NUMBER: 33,557
REFERENCE/DOCKET NUMBER: OHSU 306B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 224-6655
TELEFAX: (503) 295-6679
TELEX: 360619
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-810-134

Query Match 31.0%; Score 45; DB 4; Length 903;
Best Local Similarity 42.3%; Pred. No. 50;
Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 3 REKLRVHTGTGDEYFRFTLLADTV 28
: | | : | | : | | | | |
Db 724 KSKLFIVDQHASDEKYNFETLQAVTV 749

RESULT 11
US-08-352-902D-134
Sequence 134, Application US/08352902D
Patent No. 6191268
GENERAL INFORMATION:
APPLICANT: Liskay, Robert M.
Bronner, C. Eric
Baker, Sean M.
Bollag, Roni J.
Kolodner, Richard D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS RELATING TO DNA
MISMATCH REPAIR GENES
NUMBER OF SEQUENCES: 149
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &
Heuser
STREET: 520 S.W. Yamhill Street, Suite 200
CITY: Portland
STATE: Oregon
COUNTRY: U.S.A.
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/352,902D
FILING DATE: 09-Dec-1994
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Van Rysselberghe, Pierre C.
REGISTRATION NUMBER: 33,557
REFERENCE/DOCKET NUMBER: OHSU 306B
TELECOMMUNICATION INFORMATION:

TELEPHONE: (503) 224-6655
TELEFAX: (503) 295-6679
TELEX: 360619
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 134:
US-08-352-902D-134

Query Match 31.0%; Score 45; DB 4; Length 903;
Best Local Similarity 42.3%; Pred. No. 50;
Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 3 REKLRVHTGTGDEYFRFTLLADTV 28
: | | : | | : | | | | |
Db 724 KSKLFIVDQHASDEKYNFETLQAVTV 749

RESULT 12
US-08-513-974B-327
Sequence 327, Application US/08513974B
Patent No. 6114139
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Ohtaki, Tetsuya
APPLICANT: Fukusumi, Shoji
APPLICANT: Ohgi, Kazuhiro
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,
TITLE OF INVENTION: PRODUCTION, AND USE THEREOF
NUMBER OF SEQUENCES: 380
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/513,974B
FILING DATE: 14-SEP-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP95/01599
FILING DATE: 10-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-093989
FILING DATE: 19-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-057186
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-007177
FILING DATE: 20-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-326611
FILING DATE: 28-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-270017
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236357
FILING DATE: 30-SEP-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236356
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189274
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189273
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189272
FILING DATE: 11-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 45753
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 327:
SEQUENCE CHARACTERISTICS:
LENGTH: 129 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-513-974B-327

Query Match 30.3%; Score 44; DB 3; Length 129;
Best Local Similarity 39.3%; Pred. No. 7 6;
Matches 11; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

QY 3 REKLAVTHQTGDEYFR--FTTLADTV 28
DB 48 RNKTTTCYDTTADYLRSYFYSMCTTV 75

RESULT 13
US-08-513-974B-374
Sequence 374, Application US/08513974B
Patent No. 6114139
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Ohkaki, Tetsuya
APPLICANT: Ohgaki, Kazuhiko
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,
NUMBER OF SEQUENCES: 380
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/513.974B
FILING DATE: 14-SEP-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP95/01599
FILING DATE: 10-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-093989
FILING DATE: 19-AUG-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-057186
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-007177
FILING DATE: 20-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-326611
FILING DATE: 28-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-270017
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236357
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236356
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189274
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189273
FILING DATE: 11-AUG-1945
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189272
FILING DATE: 11-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 45753
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 374:
SEQUENCE CHARACTERISTICS:
LENGTH: 362 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-513-974B-374

Query Match 30.3%; Score 44; DB 3; Length 362;
Best Local Similarity 39.3%; Pred. No. 25;
Matches 11; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

QY 3 REKLAVTHQTGDEYFR--FTTLADTV 28
DB 185 RNKTTTCYDTTADYLRSYFYSMCTTV 212

RESULT 14
US-08-687-590-26
Sequence 26, Application US/08687590
Patent No. 6255070
GENERAL INFORMATION:
APPLICANT: Willison, Keith Robert
APPLICANT: Kubota, Hiroshi
APPLICANT: Ashworth, Alan
TITLE OF INVENTION: Folding Proteins
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,590
; FILING DATE: 31-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00192
; FILING DATE: 31-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9401791.0
; FILING DATE: 31-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9418234.2
; FILING DATE: 09-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 084619-00000005
; INFORMATION FOR SEQ. ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 545 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-687-590-26

Query Match      30.3%; Score 44; DB 4; Length 545;
Best Local Similarity 88.9%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 14 GDEYFRFT 22
Db 356 GDEYFRFT 364
```

```

RESULT 15
US-08-380-403A-6
; Sequence 6, Application US/08380403A
; Patent No. 5831024
; GENERAL INFORMATION:
; APPLICANT: MINATO, Nagahiro
; APPLICANT: HATTORI, Masakazu
; APPLICANT: HIROSHI, Kubota
; APPLICANT: MASATSUGU, Maeda
; TITLE OF INVENTION: SPA-1 PROTEIN AND GENE CODING THEREFOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/380,403A
; FILING DATE: 30-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/325,909
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-279712
; FILING DATE: 20-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-139513
; FILING DATE: 30-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
```

```

; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 53466/128/AAOK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ. ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-380-403A-6

Query Match      29.7%; Score 43; DB 2; Length 240;
Best Local Similarity 42.1%; Pred. No. 23;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
QY 10 HOTGDEYFRFTLTADTV 28
Db 11 NOGAGATFWQTLTGDDV 29
```

Search completed: January 29, 2003, 06:47:49
Job time : 13.2982 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compen Ltd.

OK protein - protein search, using sw model

Run on: January 29, 2003, 06:46:04 ; Search time 7.85965 seconds
(without alignments)
71.886 Million cell updates/sec

Title: US-09-627-165D-20
Perfect score: 145
Sequence: 1 YERERLRVTHQTGDEYFRFTLLADTV 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues
Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PC1_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PC1US_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	81.4	252	10	US-09-347-064-2
2	118	81.4	252	10	US-09-347-064-8
3	48	33.1	1562	9	US-09-860-846-35
4	48	33.1	1562	10	US-09-861-289-35
5	48	33.1	11877	9	US-09-860-846-6
6	48	33.1	11877	10	US-09-861-289-6
7	44	30.3	125	10	US-09-864-761-47339
8	44	30.3	196	12	US-10-059-964-16
9	44	30.3	559	9	US-09-854-133-396
10	44	30.3	559	10	US-09-738-973-396
11	43	29.7	278	9	US-09-738-626-5155
12	43	29.7	1029	9	US-10-033-245-22
13	43	29.7	1029	9	US-10-033-223-22
14	43	29.7	1029	9	US-10-033-167-22
15	43	29.7	1029	9	US-10-033-244-22
16	43	29.7	1029	12	US-10-033-246-22
17	43	29.7	1029	12	US-10-033-301-22
18	43	29.7	1029	12	US-10-033-326-22
19	43	29.7	3313	10	US-09-737-149-29

20	42	29.0	215	10	US-09-925-299-922	Sequence 922, App
21	42	29.0	258	9	US-09-813-453A-6	Sequence 6, Appl1
22	42	29.0	577	10	US-09-815-242-10193	Sequence 10193, A
23	41.5	28.6	335	9	US-09-810-506-2	Sequence 2, Appl1
24	41.5	28.6	516	10	US-09-740-046-4	Sequence 4, Appl1
25	41	28.3	64	10	US-09-939-980-447	Sequence 447, App
26	41	28.3	386	9	US-09-738-626-6216	Sequence 6216, Ap
27	41	28.3	593	10	US-09-815-242-11080	Sequence 11080, A
28	41	28.3	639	10	US-09-768-877-25	Sequence 25, Appl
29	41	28.3	791	9	US-10-055-364-41	Sequence 41, Appl
30	41	28.3	792	9	US-10-055-364-42	Sequence 42, Appl
31	41	28.3	4613	9	US-09-860-846-31	Sequence 31, Appl
32	41	28.3	4613	10	US-09-861-289-31	Sequence 31, Appl
33	40.5	27.9	252	9	US-09-860-670-119	Sequence 119, App
34	40.5	27.9	275	10	US-09-815-242-12069	Sequence 12069, A
35	40.5	27.9	582	10	US-09-815-242-10115	Sequence 10115, A
36	40.5	27.9	669	9	US-09-983-204-15	Sequence 15, Appl
37	40.5	27.9	669	9	US-10-133-157-4	Sequence 4, Appl1
38	40.5	27.9	1354	10	US-09-808-571A-4	Sequence 4, Appl1
39	40.5	27.9	1447	10	US-09-808-571A-2	Sequence 2, Appl1
40	40	27.6	52	10	US-09-864-761-35437	Sequence 35437, A
41	40	27.6	72	10	US-09-864-761-34449	Sequence 182, App
42	40	27.6	256	9	US-09-895-913A-182	Sequence 2, Appl1
43	40	27.6	461	9	US-10-051-186-2	Sequence 9, Appl1
44	40	27.6	464	9	US-10-051-186-9	Sequence 5, Appl1
45	40	27.6	888	10	US-09-810-796-5	

ALIGNMENTS

RESULT 1
US-09-347-064-2
Sequence 2, Application US/09347064A
Patent No. US20020045208A1
GENERAL INFORMATION:
APPLICANT: Eck, Jurgen
APPLICANT: Schmidt, Arno
APPLICANT: Zinke, Holger
TITLE OF INVENTION: Recombinant Fusion Proteins Based on
TITLE OF INVENTION: Ribosome-Inactivating Proteins of the mistletoe Viscum
FILE REFERENCE: 09282-5
CURRENT APPLICATION NUMBER: US/09/347,064A
EARLIER FILING DATE: 1999-07-02
EARLIER APPLICATION NUMBER: PCT/EP98/00009
EARLIER FILING DATE: 1998-01-02
EARLIER APPLICATION NUMBER: EP 97 10 0012.0
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 252
TYPE: PRT
ORGANISM: Viscum album
US-09-347-064-2

Query Match 81.4% Score 118. DB 10; Length 252;
Best Local Similarity 82.1% Pred. No. 3.2e-11;
Matches 23; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YERERLRVTHQTGDEYFRFTLLADTV 28
||| ||||| ||||| ||||| |||
Db 2 YERERLRVTHQTGDEYFRFTLLADTV 29

RESULT 2
US-09-347-064-8
Sequence 8, Application US/09347064A
Patent No. US20020045208A1
GENERAL INFORMATION:
APPLICANT: Eck, Jurgen
APPLICANT: Schmidt, Arno

```

; APPLICANT: Zinke, Holger
; TITLE OF INVENTION: Recombinant Fusion Proteins Based on
; TITLE OF INVENTION: Ribosome-Inactivating Proteins of the mistletoe Viscum
; TITLE OF INVENTION: album
; FILE REFERENCE: 09282-5
; CURRENT APPLICATION NUMBER: US/09/347,064A
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: PCT/EP98/00009
; EARLIER FILING DATE: 1998-01-02
; EARLIER APPLICATION NUMBER: EP 97 10 0012.0
; EARLIER FILING DATE: 1997-01-02
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Viscum album
; US-09-347-064-8

Query Match      81.4%; Score 118; DB 10; Length 252;
Best Local Similarity 82.1%; Pred. No. 3,2e-11;
Matches 23; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YERKLRVTHQTGGDEYFRFTLLADTV 28
    ||| :|||||:|||||:|||||:
Db 1 YERIRLRVTHQTGGDEYFRFTLLADTV 28

RESULT 3
US-09-860-846-35
; Sequence 35, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-860-846-35

Query Match      33.1%; Score 48; DB 9; Length 1562;
Best Local Similarity 44.8%; Pred. No. 27;
Matches 13; Conservative 5; Mismatches 9; Indels 2; Gaps 1;

QY 2 EREKL--RVTHQTGGDEYFRFTLLADTV 28
    :|| | | :| :| :| :| :| :|
Db 970 DREALARLALTGTGGFTGVSLDDLV 998

RESULT 4
US-09-861-289-35
; Sequence 35, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
```

```

; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-861-289-35

Query Match      33.1%; Score 48; DB 10; Length 1562;
Best Local Similarity 44.8%; Pred. No. 27;
Matches 13; Conservative 5; Mismatches 9; Indels 2; Gaps 1;

QY 2 EREKL--RVTHQTGGDEYFRFTLLADTV 28
    :|| | | :| :| :| :| :| :|
Db 970 DREALARLALTGTGGFTGVSLDDLV 998

RESULT 5
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-860-846-6

Query Match      33.1%; Score 48; DB 9; Length 11877;
Best Local Similarity 44.8%; Pred. No. 3e+02;
Matches 13; Conservative 5; Mismatches 9; Indels 2; Gaps 1;

QY 2 EREKL--RVTHQTGGDEYFRFTLLADTV 28
    :|| | | :| :| :| :| :| :|
Db 9658 DREALARLALTGTGGFTGVSLDDLV 9686

RESULT 6
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-861-289-6
```

Query Match 33.1%; Score 48; DB 10; Length 11877;
Best Local Similarity 44.8%; Pred. No. 3e+02;
Matches 13; Conservative 5; Mismatches 9; Indels 2; Gaps 1;
OY 2 EREKL-RVTHQTGDEYFRFTLLADTV 28
Db 9658 DREALARLTALTGTGDCFTGVSLDDLV 9686

RESULT 7

US-09-864-761-47239
; Sequence 47239, Application US/09864761
; Patent No. US20020048763a1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aegm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47239
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006115.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.58
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.49
; OTHER INFORMATION: EST_HUMAN HIT: AW873518.1, EVALU8 3.00e-19
; OTHER INFORMATION: SWISSPROT HIT: Q9U158, EVALU8 9.00e-26
US-09-864-761-47239

Query Match 30.3%; Score 44; DB 10; Length 125;
Best Local Similarity 40.9%; Pred. No. 5.8;
Matches 9; Conservative 7; Mismatches 4; Indels 2; Gaps 1;
OY 2 EREKLRVTHQTGDEYFRFTLL 23
Db 32 ETEFLKVSHPKFGD--FQYLSV 51

RESULT 8

US-10-059-964-16
; Sequence 16, Application US/10059964
; Patent No. US20020120115a1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. US20020120115a11c
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-059-964-16

Query Match 30.3%; Score 44; DB 12; Length 196;
Best Local Similarity 28.6%; Pred. No. 9.9;
Matches 8; Conservative 5; Mismatches 14; Indels 0; Gaps 0;
OY 1 YEREKLRVTHQTGDEYFRFTLLADTV 28

Db 116 WDNPALHTYTLTGSEIRNFEDILYEIT 143

RESULT 9

US-09-854-133-396
; Sequence 396, Application US/09854133
; Publication No. US20020183499a1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Monamath, Radoch
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 396
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-133-396

Query Match 30.3%; Score 44; DB 9; Length 559;
Best Local Similarity 88.9%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 14 GDEYFRFT 22

Db 370 GDEYFRFT 378

```
RESULT 10
US-09-738-973-396
; Sequence 396, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Iodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raedoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Inditias, Carol Joseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 396
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-396

Query Match      30.3%; Score 44; DB 10; Length 559;
Best Local Similarity 88.9%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      14 GDEXFFRFT 22      |||||
Db      370 GDEXFFRFT 378

RESULT 11
US-09-738-626-5155
; Sequence 5155, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIRO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5155
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5155

Query Match      29.7%; Score 43; DB 9; Length 278;
```

```
Best Local Similarity 42.9%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY      8 VTHQTDGEXFFRFTLTADYV 28      | | | | | | | |
Db      53 VVHSTMGSDYLCFRHLMWDHV 73

RESULT 12
US-10-033-245-22
; Sequence 22, Application US/10033245
; Patent No. US20020160392A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tunas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2930R1C7
; CURRENT APPLICATION NUMBER: US/10/033,245
; CURRENT FILING DATE: 2001-12-27
; PRIOR APPLICATION NUMBER: 60/095,325
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/112,851
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,145
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,511
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/115,558
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,565
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,733
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/119,341
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119,537
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/119,965
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 60/162,506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 60/170,262
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/187,202
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/28634
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
```



```
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 22
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-033-245-22

Query Match          29.7%; Score 43; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 19 RFTTLADT 27
Db 680 RFTTLADT 688

RESULT 13
US-10-033-223-22
; Sequence 22, Application US/10033223
; Patent No. US20020164646A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2930R1C9
; CURRENT APPLICATION NUMBER: US/10/033,223
; CURRENT FILING DATE: 2001-12-27
; PRIOR APPLICATION NUMBER: 60/095,325
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/112,851
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,145
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,511
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/115,558
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,565
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,733
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/119,341
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119,537
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/119,965
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 60/162,506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 60/170,262
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/187,202
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/28634
; PRIOR FILING DATE: 1999-12-01
```

```
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 22
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-033-223-22

Query Match          29.7%; Score 43; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 19 RFTTLADT 27
Db 680 RFTTLADT 688

RESULT 14
US-10-033-167-22
; Sequence 22, Application US/10033167
; Publication No. US20020182618A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2930R1C10
; CURRENT APPLICATION NUMBER: US/10/033,167
; CURRENT FILING DATE: 2001-12-27
; PRIOR APPLICATION NUMBER: 60/095,325
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/112,851
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,145
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,511
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/115,558
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,565
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,733
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/119,341
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119,537
; PRIOR FILING DATE: 1999-02-10
```

```
; PRIOR APPLICATION NUMBER: 60/119,965
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 60/162,506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 60/170,262
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/187,202
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/28634
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 22
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-033-167-22
```

```
Query Match          29.7%; Score 43; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 19 RFTTLADT 27
    |||||||
Db 680 RFTTLADT 688
```

```
RESULT 15
US-10-033-244-22
; Sequence 22, Application US/10033244
; Publication No. US2002019268A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2930R1C2
; CURRENT APPLICATION NUMBER: US/10/033,244
; PRIOR FILING DATE: 2001-12-27
; PRIOR APPLICATION NUMBER: 60/095,325
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/112,851
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,145
; PRIOR FILING DATE: 1998-12-16
```

```
; PRIOR APPLICATION NUMBER: 60/113,511
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/115,558
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,565
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,733
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/119,341
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119,537
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/119,965
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 60/162,506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 60/170,262
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/187,202
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/28634
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 22
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-033-244-22
```

```
Query Match          29.7%; Score 43; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 19 RFTTLADT 27
    |||||||
Db 680 RFTTLADT 688
```

```
Search completed: January 29, 2003, 06:50:26
Job time : 10.8596 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 06:44:49 ; Search time 13.7544 Seconds
(without alignments)
195.702 Million cell updates/sec

Title: US-09-627-165D-20

Perfect score: 145
Sequence: 1 YERKRLRVHTGTGDEYFRFTLLADTV 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being predicted,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	81.4	254	2 PD0018	mistletoe lectin I
2	50	34.5	454	2 S77005	sensory transducti
3	50	34.5	464	2 T20238	hypothetical prote
4	50	34.5	2358	2 T39569	probable alpha-glu
5	50	34.5	2371	2 T43432	alpha-glucan synth
6	49.5	34.1	1147	1 MMAX1B	myosin heavy chain
7	48.5	33.4	1168	1 MMAX1C	myosin heavy chain
8	48	33.1	174	2 T17753	hypothetical prote
9	48	33.1	318	2 AC2442	hypothetical prote
10	48	33.1	1562	2 T17411	polyketide synthas
11	47.5	32.8	666	1 E64203	ATP-dependent nucl
12	47.5	32.8	673	2 T05619	hypothetical prote
13	47	32.4	307	2 A75020	hypothetical prote
14	46	31.7	161	2 AD2540	hypothetical prote
15	46	31.7	253	2 B17547	probable tryptopha
16	46	31.7	367	2 E82974	probable transcrip
17	45.5	31.4	1100	2 T21544	hypothetical prote
18	45	31.0	354	2 B56392	beta-galactoside a
19	45	31.0	544	2 S42723	matricin - mouse
20	45	31.0	545	2 A66281	tetrahydrofolylpol
21	45	31.0	718	2 F97411	ycg4A protein (Arl
22	45	31.0	718	2 A12629	conserved hypochet
23	44.5	30.7	291	2 A72341	hypothetical prote
24	44.5	30.7	312	2 T22918	hypothetical prote
25	44.5	30.7	1113	1 A47106	myosin heavy chain
26	44	30.3	178	2 C61235	transcription antl
27	44	30.3	211	2 A60098	conserved hypochet
28	44	30.3	250	2 AB0171	probable 3'-deoxy-m
29	44	30.3	266	2 A46610	C 3.4.25.1 proteas

30	44	30.3	307	2 E71206	hypothetical prote
31	44	30.3	362	2 S33733	G protein-coupled
32	44	30.3	383	2 T36080	probable NADH dehy
33	44	30.3	398	2 S77230	hypothetical prote
34	44	30.3	544	2 A38983	TCPI ring complex
35	44	30.3	545	2 S43062	CCT (chaperonin co
36	44	30.3	545	2 S54210	chaperonin contain
37	44	30.3	769	2 S54210	outer membrane pro
38	44	30.3	838	2 T04449	hypothetical prote
39	44	30.3	848	2 G67220	heat shock protein
40	43.5	30.0	310	2 T25515	hypothetical prote
41	43.5	30.0	315	2 AFI695	transcription repr
42	43.5	30.0	350	2 C97101	biotin synthase fa
43	43.5	30.0	373	2 T35452	probable solute bi
44	43.5	30.0	448	2 B45438	myosin I beta, MMI
45	43.5	30.0	476	1 KCRTS2	stromelysin 2 (EC

ALIGNMENTS

RESULT 1

PD0018
mistletoe lectin I A chain - Viscum album (fragment)
C:Species: Viscum album
C>Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 26-Aug-1999
C:Accession: PD0018
R:Eschenburg, S.; Krauspenhaar, R.; Mikhailov, A.; Stoeva, S.; Betzel, C.; Voelter, W
Biochem. Biophys. Res. Commun. 247, 367-372, 1998
A:Title: Primary structure and molecular modeling of mistletoe lectin I from Viscum a
A:Reference number: PD0018; MUID:98308123; PMID:9642133
A:Accession: PD0018
A:Molecule type: protein
A:Residues: 1-254 <ESC>
A:Superfamily: ricin; RNA N-glycosidase homology
F;7-246/Domain: RNA N-glycosidase homology <RNC>

Query Match 81.4%; Score 118; DB 2; Length 254;
Best Local Similarity 82.1%; Pred. No. 1.5e-10;
Matches 23; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 1 YERKRLRVHTGTGDEYFRFTLLADTV 28
Db 1 YERLRLRVHTGTGDEYFRFTLLADTV 28

RESULT 2

S77005
sensory transduction histidine kinase sl10798 - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein sl10798
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 02-Sep-2000
C:Accession: S77005
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Ya
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
S:Reference number: S74322; MUID:97061201; PMID:8905221
A:Accession: S77005
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-454 <KAN>
A:Cross-references: EMBL:D64005; GB:AB001339; NID:91001779; PIDN:BA010697.1; PID:9100
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: sensory transduction histidine kinase sl10798; sensor histidine kinase
F:703-445/Domain: sensor histidine kinase homology <SHK>

Query Match 34.5%; Score 50; DB 2; Length 454;
Best Local Similarity 37.5%; Pred. No. 9.8;
Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

RESULT 7

MMAXIC

myosin heavy chain IC - Acanthamoeba castellanii
N:Contains: myosin ATPase (EC 3.6.4.1)

C:Species: Acanthamoeba castellanii
C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 19-Apr-2002

C:Accession: A33891; C34448; A24146

R:Jung, G.; Korn, E.D.; Hammer III, J.A.

Proc. Natl. Acad. Sci. U.S.A. 84, 6720-6724, 1987

A:Title: The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like and non-my

A:Reference number: A33891; MUID:88016163; PMID:3477803

A:Accession: A33891

A:Molecule type: DNA

A:Residues: 1-1168 <JUN>

A:Cross-references: GB:J02974; NID:9155624; PIDN:AAA27707.1; PID:9155625

A:Note: this gene and protein are called MIB in this paper

R:Przecka, H.; Lynch, T.J.; Martin, B.; Korn, E.D.

J. Biol. Chem. 264, 19340-19348, 1989

A:Title: The localization and sequence of the phosphorylation sites of Acanthamoeba myos

A:Reference number: A34448; MUID:90037074; PMID:2530230

A:Accession: C34448

A:Molecule type: protein

A:Residues: 308-314,'X',316-329

C:Comment: in this protein, the coiled-coil rod-like region found in many myosin heavy c

he protein is globular and does not self-associate into filaments.

C:Genetics:

A:Gene: MIC

A:Intons: 1/3: 37/3; 60/2: 100/2; 153/3; 179/3; 208/2; 242/3; 287/3; 321/3; 371/3; 428/

C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 homolo

C:Keywords: actin binding; ATP; hydrolase; nucleotide binding; P-loop; phosphoprotein; t

F:10-653/Domain: myosin motor domain homology <MMOT>

F:101-108/Region: nucleotide-binding motif A (P-loop)

F:543-564/Region: actin binding #status predicted

F:671-1168/Domain: carboxyl-terminal <CTD>

F:675-883/Region: basic

F:923-978/Region: alanine/glycine/proline-rich

F:983-1030/Domain: SH3 homology <SH3>

F:1034-1168/Region: alanine/glycine/proline-rich

F:107/Binding site: ATP (Lys) #status predicted

F:311/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match

Best Local Similarity 33.4%; Score 48.5; DB 1; Length 1168;

Matches 13; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

OY 1 YEREKLRVTHQTGDEYERF-FTTLA 25

Db 175 YLEKSRVYOTNGERNFHFYQLA 200

RESULT 8

T17753

hypothetical protein A259L - Chlorella virus PBCV-1

C:Species: Chlorella virus PBCV-1

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T17753

R:Graves, M.V.; Van Etten, J.L.

submitted to the EMBL Data Library, May 1999

A:Reference number: Z18806

A:Accession: T17753

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-174 <GRA>

A:Cross-references: EMBL:U42580; NID:94028896; PIDN:AAC96627.1

A:Experimental source: specific host Chlorella strain NC64A

C:Genetics:

A:Note: A259L

Query Match

Best Local Similarity 33.1%; Score 48; DB 2; Length 174;

Matches 8; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

OY 1 YEREKLRVTHQTGDEYERF-FTTLADTV 28

Db 5 YHHENRIRRTVGEEFMDNVTITDNL 32

RESULT 9

AC2442

hypothetical protein all15091 [Imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AC2442

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu

Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata

DNA Res. 6, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AC2442

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-318 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA876790.1; PID:917134229; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all15091

Query Match

Best Local Similarity 33.1%; Score 48; DB 2; Length 318;

Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 YEREKLRVTHQTGDEYERF 20

Db 94 FERKAOLTSQVTGPHVNF 113

RESULT 10

T17411

polyketide synthase III - Streptomyces venezuelae

C:Species: Streptomyces venezuelae

C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000

C:Accession: T17411

R:Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.

Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998

A:Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezue

A:Reference number: Z18773; MUID:98445333; PMID:9770448

A:Accession: T17411

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1562 <XUE>

A:Cross-references: EMBL:AF079138; NID:93808326; PID:93800836; PIDN:AAC69331.1

C:Genetics:

A:Gene: p1X1111

C:Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-ox

homology; [acyl-carrier-protein] S-malonyltransferease homology

C:Keywords: antibiotic biosynthesis; carrier protein

F:1403-1474/Domain: acyl carrier protein homology <AC>

Query Match

Best Local Similarity 33.1%; Score 48; DB 2; Length 1562;

Matches 13; Conservative 5; Mismatches 9; Indels 2; Gaps 1;

OY 2 EREKLRVTHQTGDEYERF-FTTLADTV 28

Db 970 DRELAARLTRALTTGDEFTGVSLDDLV 998

RESULT 11

E64203

ATP-dependent nuclease addA homolog - Mycoplasma genitalium

C:Species: Mycoplasma genitalium

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000

C:Accession: E64203

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.

M.; Fuhmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,

, C.A.: Venter, J.C.
 Science 270, 397-403, 1995
 A:Title: The minimal gene complement of Mycoplasma genitalium.
 A:Reference number: A64200; MUID:96026346; PMID:756993
 A:Accession: E64203
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1666 <11GR>
 A:Cross-references: GB:039682; GB:LA3967; NID:93844634; PIDN:AAC71248.1; PID:91045703; T
 A:Experimental source: strain G-37
 A:Genetics:
 A:Genetic code: SGC3
 C:Superfamily: ATP-dependent nuclease adda homolog

Query Match 32.8%; Score 47.5; DB 1; Length 666;
 Best Local Similarity 57.9%; Pred. No. 37;
 Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

OY 5 KLRVTHQTGD-EYEFRT 22
 |||||: ||:| |
 Db 576 KLRVTHQTGDVYRRLHT 594

RESULT 12
 T05619
 hypothetical protein F20D10.40 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
 C:Accession: T05619
 R:Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, R.F.
 submitted to the Protein Sequence Database, February 1999
 A:Reference number: Z15420
 A:Accession: T05619
 A:Molecule type: DNA
 A:Residues: 1-673 <BEV>
 A:Cross-references: EMBL:AL035538
 A:Experimental source: cultivar Columbia; BAC clone F20D10
 C:Genetics:
 A:Map position: 4
 A:Introns: 239/1; 308/1; 396/3; 452/1; 533/3; 596/2
 A:Note: F20D10.40

Query Match 32.8%; Score 47.5; DB 2; Length 673;
 Best Local Similarity 56.2%; Pred. No. 37;
 Matches 9; Conservative 6; Mismatches 0; Indels 1; Gaps 1;

OY 13 TGDYFRFTLLADTV 28
 ||:||||:|:
 Db 430 TGDYFRFTLLADTV 444

RESULT 13
 A75020
 hypothetical protein PAB110 - Pyrococcus abyssi (strain Orsay)
 C:Species: Pyrococcus abyssi
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: A75020
 R:anonymous, Genoscope
 submitted to the EMBL Data Library, July 1999
 A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
 A:Reference number: A75001
 A:Accession: A75020
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-307 <KAW>
 A:Cross-references: GB:AJ248288; GB:AL096836; NID:95458960; PIDN:CAB50599.1; PID:9545911
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: PAB110

Query Match 32.4%; Score 47; DB 2; Length 307;
 Best Local Similarity 52.6%; Pred. No. 18;
 Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 YEREKLRVTHQTGDYFR 19
 |||||:|:|:
 Db 20 YPESLRKISOREGLEYR 38

RESULT 14
 AD2540
 hypothetical protein al17607 [imported] - Nostoc sp. (strain PCC 7120) plasmid pcc712
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AD2540
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
 Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
 DNA Res. 6, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD2540
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-161 <KUR>
 A:Cross-references: GB:AP003602; PIDN:BA877250.1; PID:917134692; GSPDB:GN00181
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Genome: plasmid

Query Match 31.7%; Score 46; DB 2; Length 161;
 Best Local Similarity 34.8%; Pred. No. 13;
 Matches 8; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

OY 2 EREKLRVTHQTGDYFRFTLL 24
 ||:|:||||:|:
 Db 122 EMNNVAKTAVTTGDKRYKSTIL 144

RESULT 15
 B71547
 probable tryptophane synthase (alpha chain) - Chlamydia trachomatis (serotype D, stra
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: B71547
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
 A:Reference number: A71570; MUID:9900809; PMID:9784136
 A:Accession: B71547
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-253 <ARN>
 A:Cross-references: GB:AE001291; GB:AE001273; NID:93328573; PIDN:AAC67762.1; PID:9332
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: trpA

Query Match 31.7%; Score 46; DB 2; Length 253;
 Best Local Similarity 45.8%; Pred. No. 21;
 Matches 11; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

OY 5 KLRVTHQTGDYFRFTLLADTV 28
 |||||:|:|:
 Db 227 KTRAFVHTTWDSSVETTLAQT 250

Search completed: January 29, 2003, 06:47:18
 Job time : 15.7544 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 06:44:49 ; Search time 18.1754 Seconds
(without alignments)
63.896 Million cell updates/sec

Title: US-09-627-165D-20

Perfect score: 145
Sequence: 1 YEREKRVTHQGTGDEXFRFTLLADTV 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	81.4	254	1	MLA_VISAL
2	50	34.5	2358	1	MOKD_SCHPO
3	49.5	34.1	1147	1	MYSH_ACACA
4	48.5	33.4	1168	1	MYSC_ACACA
5	47.5	32.8	666	1	Y032_MYCGE
6	46	31.7	253	1	TRPA_CHLTR
7	45	31.0	354	1	FUT2_RABIT
8	45	31.0	587	1	FOIC_HUMAN
9	45	31.0	1042	1	SPAI_HUMAN
10	44.5	30.7	1113	1	MYSD_DICDI
11	44	30.3	178	1	MUSG_NEMIA
12	44	30.3	250	1	KDSB_YERPE
13	44	30.3	266	1	PSB4_YENST
14	44	30.3	362	1	P2YR_CHICK
15	44	30.3	362	1	P2YR_MELGA
16	44	30.3	544	1	TCPC_HUMAN
17	44	30.3	545	1	TCPC_MOUSE
18	44	30.3	547	1	TCPC_XENIA
19	44	30.3	1037	1	SPAI_MOUSE
20	43.5	30.0	414	1	CDVI_MOUSE
21	43.5	30.0	455	1	PEX3_PICPA
22	43.5	30.0	476	1	MM10_MOUSE
23	43.5	30.0	476	1	MM10_RAT
24	43.5	30.0	1028	1	MYIC_HUMAN
25	43.5	30.0	1028	1	MYIC_MOUSE
26	43	29.7	261	1	T2S9_STRAU
27	43	29.7	496	1	DPOW_MOUSE
28	43	29.7	544	1	TCPC_MOUSE
29	43	29.7	714	1	GPK2_DROME
30	43	29.7	3301	1	CLR3_MOUSE
31	43	29.7	3312	1	CLR3_HUMAN
32	43	29.7	3312	1	CLR3_RAT
33	43	29.7	5035	1	RYR1_PIG

34	42.5	29.3	87	1	R31C_ECO57	08x9t8 escherichia
35	42.5	29.3	510	1	FOIC_CAMEL	009509 caenorhabd
36	42.5	29.3	1986	1	W4EMENI	003149 emericella
37	42	29.0	181	1	Y4WG_RHISN	P55685 rhizobium s
38	42	29.0	201	1	A1AG_HUMAN	P02763 homo sapien
39	42	29.0	373	1	P2YR_BOVIN	P48042 bos taurus
40	42	29.0	373	1	P2YR_HUMAN	P47900 homo sapien
41	42	29.0	577	1	SYR_ECO57	08xh2 escherichia
42	42	29.0	577	1	SYR_ECOLI	P1875 escherichia
43	42	29.0	649	1	FLR3_HUMAN	09uz0 homo sapien
44	42	29.0	701	1	VATI_SULSO	09uw3 sulfobus
45	42	29.0	1157	1	C8AA_BACUK	Q45704 bacillus th

ALIGNMENTS

RESULT 1	ID	MLA_VISAL	STANDARD	PRT	254 AA
AC	P81446				
DT	15-DEC-1998	(Rel. 37, Created)			
DT	15-DEC-1998	(Rel. 37, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Beta-galactoside specific lectin I A chain (MLA) (ML-I A) (rRNA N-glycosidase) (EC 3.2.2.22).				
DE	glycosidase (EC 3.2.2.22).				
OS	Viscum album (European mistletoe).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
OC	Santalales; Viscaceae; Viscum.				
OX	NCBI_TaxID=3972;				
RN	[1]				
RP	SEQUENCE.				
RC	STRAIN-Subsp. album.				
RX	MEDLINE=97134581; PubMed=8930141;				
RA	Huguet Soler M., Stoeva S., Schwaborn C., Wilhelm S., Stiefel T.,				
RA	Voelter W.;				
RT	"Complete amino acid sequence of the A chain of mistletoe lectin I.";				
RL	FEBS Lett. 399:153-157(1996).				
CC	-I- FUNCTION: THE A CHAIN IS RESPONSIBLE FOR INHIBITING PROTEIN SYNTHESIS THROUGH THE CATALYTIC INACTIVATION OF 60S RIBOSOMAL SUBUNITS BY REMOVING ADENINE FROM POSITION 4,324 OF 28 S RNA. THE B CHAIN BINDS TO CELL RECEPTORS AND PROBABLY FACILITATES THE ENTRY INTO THE CELL OF THE A CHAIN; B CHAINS ARE ALSO RESPONSIBLE FOR CELL AGGLUTINATION (LECTIN ACTIVITY).				
CC	-I- CATALYTIC ACTIVITY: Endoglycosylase of the N-glycosidic bond at one specific adenosine on the 28S rRNA.				
CC	-I- SUBUNIT: DISULFIDE-LINKED DIMER OF A AND B CHAINS.				
CC	-I- MISCELLANEOUS: TWO ISOFORMS OF MLA EXIST: GLYCOSYLATED FORM MLA AND NON-GLYCOSYLATED FORM MLA.				
CC	-I- SIMILARITY: BELONGS TO THE RIBOSOME-INACTIVATING PROTEIN FAMILY. TYPE 2 RIP SUBFAMILY.				
CC	HSSP: P11140: 1ABR.				
DR	InterPro: IPR001574; RIP.				
DR	Pfam: PF00161; RIP; 1.				
DR	PRINTS: PR00396; SHIGARICIN.				
DR	PROSITE: PS00275; SHIGA_RICIN; FALSE_NEG.				
KW	Plant defense; Hydrolase; Toxin; Repeat; Glycoprotein; Lectin.				
FT	ACT_SITE	165	165		
FT	CARBOHYD	112	112		
FT	VARIANT	15	15		
FT	VARIANT	66	66		
FT	VARIANT	112	112		
FT	VARIANT	116	116		
FT	VARIANT	133	134		
FT	VARIANT	140	140		
FT	VARIANT	144	144		
FT	VARIANT	151	151		
FT	VARIANT	179	179		
FT	VARIANT	184	184		
FT	VARIANT	190	190		
FT	VARIANT	218	218		
FT	VARIANT	223	224		
FT	VARIANT				


```

DR ProDom; PD000065; SH3; 1.
DR SMART; SM00355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Myosin; ATP-binding; Phosphorylation; Multigene family; SH3 domain.
FT DOMAIN 1 670 MYOSIN HEAD-LIKE.
FT DOMAIN 671 907 TAIL HOMOLOG REGION 1 (TH.1).
FT DOMAIN 908 1089 GLY/PRO/ALA-RICH (TH.2).
FT DOMAIN 1090 1147 SH3.
FT NP_BIND 103 110 ATP (POTENTIAL).
SQ SEQUENCE 1147 AA; 124958 MW; B76DE9C076381054 CRC64;

Query Match 34.1%; Score 49.5; DB 1; Length 1147;
Best Local Similarity 53.8%; Pred. No. 13;
Matches 14; Conservative 1; Mismatches 10; Indels 1; Gaps 1;

Oy 1 YEREKLRVTHQTTGDEYFR-FITILLA 25
Db 179 YLEKSRVTFQTRGERSFHFYQLLA 204

RESULT 4
MYSC_ACACA STANDARD; PRT; 1168 AA.
AC P10569;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin IC heavy chain.
GN MIC.
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxId=5735;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88016163; PubMed=3477803;
RA Jung G., Korn E.D., Hammer J.A. III;
RT "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
RL and non-myosin-like sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86259656; PubMed=3014500;
RA Hammer J.A. III, Jung G., Korn E.D.;
RT "Genetic evidence that Acanthamoeba myosin I is a true myosin."
RL Proc. Natl. Acad. Sci. U.S.A. 83:4655-4659(1986).
RN [3]
RP PHOSPHORYLATION SITE.
RX MEDLINE=90037074; PubMed=2530230;
RA Brzeska H., Lynch T.J., Martin B., Korn E.D.;
RT "The localization and sequence of the phosphorylation sites of
RL Acanthamoeba myosin I. An improved method for locating the
RT phosphorylated amino acid."
RL J. Biol. Chem. 264:19340-19348(1989).
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -1- DOMAIN: TH.1 BINDS DIRECTLY TO ANTONIC PHOSPHOLIPID MEMBRANES;
CC MYOSIN I CAN THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND VICE
CC VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER WITH
CC THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE MOLECULES
CC OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
CC -1- MISCELLANEOUS: THIS ORGANISM EXPRESSES AT LEAST THREE ISOFORMS OF
CC MYOSIN I HEAVY-CHAIN, ENCODED BY GENES MIA, MIB, AND MIC.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE MYOSIN IB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02974; AAA27707.1; -
DR PIR; A33891; MMAXIC.
DR HSSP; P08799; 1MND.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000065; SH3; 1.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Myosin; ATP-binding; Phosphorylation; Multigene family; SH3 domain.
FT DOMAIN 1 670 MYOSIN HEAD-LIKE.
FT DOMAIN 671 922 TAIL HOMOLOG REGION 1 (TH.1).
FT DOMAIN 923 975 GLY/PRO/ALA-RICH (TH.2).
FT DOMAIN 976 1035 SH3.
FT DOMAIN 1036 1168 GLY/PRO/ALA-RICH (TH.2).
FT NP_BIND 101 108 ATP (POTENTIAL).
FT MOD_RES 311 311 PHOSPHORYLATION.
SQ SEQUENCE 1168 AA; 127309 MW; D07084B373A3732 CRC64;

Query Match 33.4%; Score 48.5; DB 1; Length 1168;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 13; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

Oy 1 YEREKLRVTHQTTGDEYFR-FITILLA 25
Db 175 YLEKSRVYQTGERNFHFYQLLA 200

RESULT 5
Y032_MYCGE STANDARD; PRT; 666 AA.
ID Y032_MYCGE
AC P47278;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG032.
GN MG032.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxId=2097;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569393;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischman R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lueder T.S.,
RA Peterson S.N., Smith H.O., Hutchison G.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -1- SIMILARITY: BELONGS TO THE MG032 / MG096 / MG288 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39682; AAC71248.1; -

```

DR TIGR: MG032: -
 DR InterPro: IPR004306; MG032/096/288_1.
 DR InterPro: IPR004319; MG032/096/288_2.
 DR Pfam: PF03072; DUF237; 1.
 DR Pfam: PF03086; DUF240; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 666 AA; 77302 MW; D703C107A8E2AB7F CRC64;

Query Match 32.8%; Score 47.5; DB 1; Length 666;
 Best Local Similarity 57.9%; Pred. No. 15;
 Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

OY 5 KLRVHTQGTGD-EYFRFT 22
 ||||| 11 11 1
 Db 576 KLRVHTQGTGDVYRRLHT 594

RESULT 6

TRPA_CHLTR STANDARD; PRT; 253 AA.
 ID TRPA_CHLTR
 AC 084173; 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tryptophan synthase alpha chain (EC 4.2.1.20).
 GN TRPA OR CPT17.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/Cx;
 RX MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kalmann S., Lammell C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis.";
 RL Science 283:754-759(1998).
 CC -1- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOLE CLEAVAGE
 OF INDOLEGLYCEROL PHOSPHATE TO INDOLE AND GLYCERALDEHYDE 3-
 PHOSPHATE.
 CC -1- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
 = L-tryptophan + glycerolaldehyde 3-phosphate + H(2)O.
 CC -1- PATHWAY: tryptophan biosynthesis; fifth (last) step.
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS (BY
 SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TRPA FAMILY.
 CC -1- CAUTION: THIS TRPA IS HIGHLY DIVERGENT COMPARED TO OTHER BACTERIAL
 TRPA. AS C. TRACHOMATIS SEEMS TO HAVE LOST PART OF THE TRP
 BIOSYNTHETIC OPERON, IT IS POSSIBLE THAT THIS PROTEIN IS NOT
 ACTIVE.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: AE001291; AC67762.1; -
 DR HSSP: P00929; 2TYS.
 DR PHCT_2DPAGE: 084173; -
 DR InterPro: IPR002028; TRP_synthaseA.
 DR Pfam: PF00290; TRP_synA; 1.
 DR PRODOM: PD001535; TRP_synthaseA; 1.
 DR TIGRPFAM: TIGR00262; trpa; 1.
 DR PROSITE: PS00167; TRP_SYNTHASE_ALPHA; FALSE NEG.
 KW Tryptophan biosynthesis; Lyase; Complete proteome.
 SQ SEQUENCE 253 AA; 28056 MW; DIA704445C630AEA CRC64;

Query Match 31.7%; Score 46; DB 1; Length 253;

Best Local Similarity 45.8%; Pred. No. 8.5;
 Matches 11; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

OY 5 KLRVHTQGTGD-EYFRFTLTADTV 28
 ||||| 11 11 11
 Db 227 KLRVHTQGTGDVYRRLHT 250

RESULT 7

FUT2_RABIT STANDARD; PRT; 354 AA.
 ID FUT2_RABIT
 AC 010983;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Galactoside 2-L-fucosyltransferase 2 (EC 2.4.1.69) (Secretor blood
 group alpha-2-fucosyltransferase) (GDP-L-fucose:beta-D-
 galactoside 2-alpha-L-fucosyltransferase 2) (Alpha(1,2)Ft 2)
 DE (Fucosyltransferase 2).
 GN FUT2 OR SECT OR RFT-II.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95238380; PubMed=7721792;
 RA Hirotsu S., Kusunoki S., Kanazawa I., Tsuji S.;
 RT "Molecular cloning and expression of two types of rabbit beta-
 galactoside alpha 1,2-fucosyltransferase.";
 RL J. Biol. Chem. 270:8844-8850(1995).
 CC -1- FUNCTION: CREATES A SOLUBLE PRECURSOR OLIGOSACCHARIDE FUC-ALPHA
 (1,2)GALBETA-) CALLED THE H ANTIGEN WHICH IS AN ESSENTIAL
 SUBSTRATE FOR THE FINAL STEP IN THE SOLUBLE A AND B ANTIGEN
 SYNTHESIS PATHWAY. H AND SE ENZYMS FUCOSYLATE THE SAME ACCEPTOR
 SUBSTRATES BUT EXHIBIT DIFFERENT KM VALUES.
 CC -1- CATALYTIC ACTIVITY: GDP-L-fucose + beta-D-galactosyl-R = GDP +
 alpha-L-fucosyl-1,2-beta-D-galactosyl-R.
 CC -1- PATHWAY: glycosylation.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
 FORM IN TRANS CISTERNAE OF GOLGI.
 CC -1- TISSUE SPECIFICITY: SALIVARY AND LACTATING MAMMARY GLANDS.
 CC -1- MISCELLANEOUS: THERE ARE THREE GENES (FUT1, FUT2 AND FUT3) WHICH
 ENCODE GALACTOSIDE 2-L-FUCOSYLTRANSFERASE IN RABBIT. THEY ARE
 EXPRESSED IN A TISSUE-SPECIFIC MANNER.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 11.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: X80225; CA56512.1; -
 DR InterPro: IPR002516; GF_11.
 DR Pfam: PF01531; Glyco_Transf_11; 1.
 KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
 FT Signal-anchor; Golgi stack.
 FT DOMAIN 1 22
 FT TRANSMEM 23 43
 FT DOMAIN 44 354
 FT TRANSMEM 197 197
 FT CARBOHYD 291 291
 FT CARBOHYD 317 317
 FT SEQUENCE 354 AA; 40035 MW; 1E2B831F9DA6CCB4 CRC64;

Query Match 31.0%; Score 45; DB 1; Length 354;
 Best Local Similarity 37.5%; Pred. No. 18;
 Matches 9; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

```

0Y      5 KLRTVHOTGDVFRFTTLADTV 28
       :| | | | | | | | | |
Db      17 RLRAHPSTVSTYFLFTFVSTV 40

RESULT 8
FOLC_HUMAN
ID      FOLC_HUMAN      STANDARD;      PRT;      587 AA.
AC      005932;
DT      01-JUN-1994 (Rel. 29, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Poly(glycylglutamate synthase, mitochondrial precursor (EC 6.3.2.17)
GN      Poly(glycyl-gamma-glutamate synthetase) (PF65).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC      NCBI_TaxID=9606;
[1]
RP      SEQUENCE OF 1-353 FROM N.A.
RA      Chen L., Qi H., Korenberg J., Shane B.;
RL      Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RP      SEQUENCE OF 20-587 FROM N.A.
RC      TISSUE=Lymphocytes;
RX      MEDLINE=93028422; PubMed=1409616;
RT      Garrow T.A., Admon A., Shane B.;
RT      "Expression cloning of a human cDNA encoding
RT      folylpoly(gamma-glutamate) synthetase and determination of its
RT      primary structure.";
RL      Proc. Natl. Acad. Sci. U.S.A. 89:9151-9155(1992).
[3]
RP      SEQUENCE OF 1-107 FROM N.A., AND ALTERNATIVE INITIATION.
RC      TISSUE=Placenta;
RX      MEDLINE=95238480; PubMed=7721888;
RA      Freemanlle S.J., Taylor S.M., Krystal G., Moran R.G.;
RT      "Upstream organization of and multiple transcripts from the human
RT      folylpoly-gamma-glutamate synthetase gene.";
RL      J. Biol. Chem. 270:9579-9584(1995).
[4]
RP      SEQUENCE OF 102-587 FROM N.A.
RC      TISSUE=Placenta;
RX      MEDLINE=96105015; PubMed=8521387;
RA      Taylor S.M., Freemanlle S.J., Moran R.G.;
RT      "Structural organization of the human folylpoly-gamma-glutamate
RT      synthetase gene: evidence for a single genomic locus.";
RL      Cancer Res. 55:6030-6034(1995).
[5]
RN      CHARACTERIZATION.
RP      MEDLINE=87157665; PubMed=3828320;
RA      Cichowicz D.J., Shane B.;
RT      "Mammalian folylpoly-gamma-glutamate synthetase. 1. Purification and
RT      general properties of the hog liver enzyme.";
RL      Biochemistry 26:504-512(1987).
-1- FUDCTION: CONVERSION OF FOLATES TO POLYGLUTAMATE DERIVATIVES. THIS
ALLOWS TISSUES TO CONCENTRATE FOLATE AT HIGHER LEVELS THAN IN
PLASMA.
-1- CATALYTIC ACTIVITY: ATP + (tetrahydrofolyl-1-[Glu])(N) + L-glutamate
= ADP + phosphate + (tetrahydrofolyl-1-[Glu])(N+1).
-1- SUBUNIT: MONOMER.
-1- SUBCELLULAR LOCATION: Mitochondrial and cytoplasmic.
-1- ALTERNATIVE PRODUCTS: 2 isoforms; mitochondrial (shown here) and
cytoplasmic; are produced by alternative initiation.
-1- SIMILARITY: BELONGS TO THE POLY(POLYGLUTAMATE SYNTHASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/)

```

```

CC      or send an email to license@lsb-sib.ch.
-----
DR      EMBL: M98045; AAA35852.1; ALT_INIT.
DR      EMBL: U14939; AAA85815.1; -.
DR      EMBL: U40868; AAA87568.1; -.
DR      EMBL: U40863; AAA87568.1; JOINED.
DR      EMBL: U40864; AAA87568.1; JOINED.
DR      EMBL: U40865; AAA87568.1; JOINED.
DR      EMBL: U40866; AAA87568.1; JOINED.
DR      EMBL: U40867; AAA87568.1; JOINED.
DR      EMBL: U24253; AAC13871.1; -.
DR      EMBL: U24252; AAC13871.1; JOINED.
DR      PIR: A46281; A46281.
DR      HSSP: p15925; 1FGS.
DR      Genew: HGNC:3824; PF03.
DR      MIM: 136510; -.
DR      InterPro: IPR001645; Polylgl_synthase.
DR      InterPro: IPR004101; Mur_ligase.C.
DR      Pfam: PF02875; Mur_ligase.C. 1.
DR      PROSITE: PS01011; FOLYPOLYGLU_SYNT_1; 1.
DR      PROSITE: PS01012; FOLYPOLYGLU_SYNT_2; 1.
KW      Ligase; One-carbon metabolism; ATP-binding; Mitochondrion;
KW      Transic peptide; Alternative initiation.
FT      TRANSIT 1 42
FT      CHAIN 43 587
FT      CHAIN 43 587
FT      CHAIN 43 587
FT      INIT_MER 43 43
FT      NP_BIND 103 109
FT      CONFLICT 22 22
FT      CONFLICT 22 22
SQ      SEQUENCE 587 AA; 64609 MW; 5AF81409F5F77E5C CRC64;

Query Match
Best Local Similarity 31.0%; Score 45; DB 1; Length 587;
Matches 12; Conservative 5; Mismatches 6; Indels 26; Gaps 1.

OY      3 REKLRVTHQTTGDE-----YRFITLTA 25
DB      137 RRRIRINGPISPELFTKYFWRLYHRLLEPTKGGSCVMPYRFRPLITMA 185
      ||:::|:-|
      ||:::|:-|

RESULT 9
SPAL_HUMAN STANDARD: PRT; 1042 AA.
AC Q96FS4; Q14518; O60618; O60484;
DT 15-JUN-2002 (Rel. 41; Created)
DT 15-JUN-2002 (Rel. 41; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE Signal-induced proliferation-associated protein 1 (Sipa-1) (GTPase-
DE activating protein Sipa-1) (p130 Sipa-1).
GN Sipa1 OR SPAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RX NCBI_TaxID=9606;
RA SEQUENCE FROM N.A. AND FUNCTION.
RA TISSUE-Peripheral blood lymphocytes;
RX MEDLINE=98010656; PubMed=9346962;
RA Kurachi H., Wada Y., Tsukamoto N., Maeda M., Kubota H., Hattori M.,
RA Iwai K., Minato N.,
RT "Human Sipa-1 product selectively expressed in lymphoid tissues is a
RT specific GTPase-activating protein for Rap1 and Rap2."
RL J. Biol. Chem. 272:28081-28088(1997).
RL (2)
RA SEQUENCE FROM N.A.
RX MEDLINE=98322269; PubMed=9651531;
RA Ebrahimi S., Wang E., Udar N., Arnold E., Burbee D., Small K.,
RA Savicki M.P.;
RT "Genomic organization and cloning of the human homologue of murine
RT Sipa-1."
RL Gene 214:215-221(1998).

```


Db 181 YLEKSRVYQTKGERNFHIFYOLLS 206

RESULT 11

ID NUSG_NEIMA STANDARD: PRT: 178 AA.

AC Q9JRD9; 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription antitermiation protein nusg.
 GN NUSG OR NMA0147 OR NMB0126.
 OS Neisseria meningitidis (serogroup A), and
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxId=65699, 491;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrrell B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis Z2491."
 RL Nature 404:502-506(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RX MEDLINE=2015755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eelsen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwinn M.L., Deboy R., Petersen J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Clechco A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
 RA Cotton M.D., Ullrich T.R., Khouli H., Qin H., Vamathevan J.,
 RA Gill J., Scarlato V., Masignani V., Pizsa M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.,
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58."
 RL Science 287:1809-1815(2000).

-1- FUNCTION: INFLUENCES TRANSCRIPTION TERMINATION AND
 ANTITERMINATION. ACTS AS A COMPONENT OF THE TRANSCRIPTION COMPLEX,
 AND INTERACTS WITH THE TERMINATION FACTOR RHO AND RNA POLYMERASE
 (BY SIMILARITY).
 -1- SIMILARITY: BELONGS TO THE NUSG FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: A162752; CAB83462.1; -;
 DR EMBL: AE002371; AAF40585.1; -;
 DR TIGR: NMB0126; -;
 DR InterPro: IPR000302; K0W_motif.
 DR InterPro: IPR001062; NUSG.
 DR Pfam: PF00467; K0W; 1.
 DR Pfam: PF02357; NUSG; 1.
 DR PRINTS: PR00338; NUSGTNSCPCT.
 DR TIGRPFAMs: TIGR00922; nusg; 1.
 DR PROSITE: PS01014; NUSG; 1.
 KW Transcription termination; Complete proteome.
 SQ SEQUENCE 178 AA; 20550 MW; 3171FDDC957EFCB3 CRC64;

Query Match 30.3%; Score 44; DB 1; Length 178;
 Best Local Similarity 62.5%; Pred. No. 12;

Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 YEREKLRVTHQTGDE 16
 ||| ||||| | | |
 Db 149 YEREKLRVSVQIFGRE 164

RESULT 12

ID KDSB_YERPE STANDARD: PRT: 250 AA.

AC 08ZGA4; 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 3-deoxy-manno-octulosonate cytidyltransferase (EC 2.7.7.38) (CMP-KDO
 synthetase) (CMP-2-keto-3-deoxyoctulosonic acid synthetase) (CKS).
 GN KDSB OR YP01400.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCBI_TaxId=632;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Bivovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague."
 RL Nature 413:523-527(2001).

-1- FUNCTION: Activates KDO (a required 8-carbon sugar) for
 incorporation into bacterial lipopolysaccharide in gram-negative
 bacteria (By similarity).
 CC -1- CATALYTIC ACTIVITY: CTP + 3-deoxy-D-manno-octulosonate =
 CC diphosphate + CMP-3-deoxy-D-manno-octulosonate.
 CC -1- PATHWAY: Liposaccharide biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE KDSB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AJ41418; CAC90229.1; -;
 DR InterPro: IPR003529; Cytidylyl-trans.
 DR InterPro: IPR004528; KDSB.
 DR Pfam: PF02348; Cytidylyl-trans; 1.
 DR TIGRPFAMs: TIGR00466; kdsb; 1.
 KW Lipopolysaccharide biosynthesis; Transferase; Nucleotidyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 250 AA; 27336 MW; 5D29AF9F07CA4F08 CRC64;

Query Match 30.3%; Score 44; DB 1; Length 250;
 Best Local Similarity 36.0%; Pred. No. 17;
 Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

OY 1 YEREKLRVTHQTGDEYRFITLLA 25
 :|||: :| | | : | : |
 Db 162 WEREFAQSKETIGDCELRHIGIYA 186

RESULT 13

ID PSB4_YEAST STANDARD: PRT: 266 AA.
 AC P30657;

DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Proteasome component PRE4 (EC 3.4.25.1) (Macropain subunit PRE4)
 DE (Proteinase 15CE subunit PRE4) (Multicatalytic endopeptidase complex
 DE subunit PRE4).
 GN PRE4 OR YFR050C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NC NCBI_TaxID:4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93155198; PubMed-8381431;
 RA Hilt W., Enekel C., Grubler A., Singer T., Wolf D.H.;
 RT "The PRE4 gene codes for a subunit of the yeast proteasome necessary
 RT for peptidylglutamyl-peptide-hydrolyzing activity. Mutations link the
 RT proteasome to stress- and ubiquitin-dependent proteolysis.";
 RL J. Biol. Chem. 268:3479-3486(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RX MEDLINE-95400292; PubMed-7670463;
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
 RA Sasakura S.-I., Sasakura M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamazaki M., Tashiro H., Eki T.;
 RT "Analysis of the nucleotide sequence of chromosome VI from
 RT Saccharomyces cerevisiae.";
 RL Nat. Genet. 10:261-268(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RX MEDLINE-96287652; PubMed-8666379;
 RA Eki T., Naitou M., Hagiwara H., Ozawa M., Sasakura S.-I.,
 RA Sasakura M., Tsuchiya Y., Shibata T., Hanaoka F., Murakami Y.;
 RT "Analysis of a 36.2 kb DNA sequence including the right telomere of
 RT chromosome VI from Saccharomyces cerevisiae.";
 RL Yeast 12:149-167(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 34-266.
 RX MEDLINE-97242404; PubMed-9087403;
 RA Groll M., Ditzel L., Lowe J., Stock D., Boehlter M., Bartunik H.D.,
 RA Huber R.;
 RT "Structure of 20S proteasome from yeast at 2.4-A resolution.";
 RL Nature 386:463-471(1997).
 CC -1- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
 CC WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ANG,
 CC PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR
 CC SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC
 CC ACTIVITY. PRE3 AND PRE4 ARE NECESSARY FOR THE PEPTIDYL-GLUTAMYL-
 CC PEPTIDE-HYDROLYZING ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad
 CC specificity.
 CC -1- SUBUNIT: YEAST PROTEASOME SEEMS TO BE COMPOSED OF 14 DIFFERENT
 CC SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-31 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X68663; CAA48629.1; -
 CC EMBL: D50617; BAA09289.1; -
 CC PIR: S29703; S29703.
 CC PIR: A46610; A46610.
 CC PDB: 1RYP; 15-APR-98.
 CC MEROPS: T01.987; -

DR SCD; S0001946; PRE4.
 DR InterPro: IPR000243; Proteasome_B.
 DR InterPro: IPR001353; Protsme_protease.
 DR Pfam: PF00227; Proteasome; 1.
 DR PROSITE: PS00854; PROTEASOME_B; 1.
 KW Proteasome; Hydrolyase; Protease; 3d-structure.
 SQ SEQUENCE 266 AA; 29443 MW; E585BB3B5D0C8E2C CRO64;
 Query Match 30.3%; Score 44; DB 1; Length 266;
 Best Local Similarity 41.2%; Pred. No. 18;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 Qy 11 OTTGDEYEFRTLLADT 27
 Db 153 QSNGDQFLRYNLGVT 169
 RESULT 14
 ID P2YR_CHICK STANDARD; PRT; 362 AA.
 AC P34996;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE P2Y purinoceptor 1 (ATP receptor) (P2Y1) (Purinergic receptor).
 GN P2RY1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE-9328340; PubMed-8508924;
 RA Webb T.E., Simon J., Krishnak B.J., Bateson A.N., Smart T.G.,
 RA King B.F., Burnstock G., Barnard E.A.;
 RT "Cloning and functional expression of a brain G-protein-coupled ATP
 RT receptor.";
 RL FEBS Lett. 324:219-225(1993).
 RN [2]
 RP 3D-STRUCTURE MODELING.
 RX MEDLINE-97026278; PubMed-8672457;
 RA van Rhee A.M., Fischer B., van Galen P.J.M., Jacobson K.A.;
 RT "Modelling the p2y purinoceptor using rhodopsin as template.";
 RL Drug Des. Discov. 13:133-140(1995).
 CC -1- FUNCTION: RECEPTOR FOR EXTRACELLULAR ADENINE NUCLEOTIDES SUCH AS
 CC ATP AND ADP. SEEMS TO MEDIATE ITS ACTION VIA A PERTUSSIS TOXIN
 CC INSENSITIVE G-PROTEIN, PROBABLY BELONGING TO THE GQ FAMILY THAT
 CC ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: BRAIN, SPINAL CORD, GASTROINTESTINAL TRACT,
 CC SPLEEN AND LEG MUSCLE. IS NOT DETECTED IN THE HEART, LIVER,
 CC STOMACH, LUNG AND KIDNEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X73268; CAA51716.1; -
 CC PIR: S33733; S33733.
 CC PDB: 1DD; 11-JUL-96.
 CC InterPro: IPR000276; GPCR_Rhodopsn.
 CC Pfam: PF00001; 7tm_1; 1.
 CC PRINTS: PR00237; GPCR_Rhodopsn.
 CC PROSITE: PS00237; G_PROTEIN_RECEPTOR_FL_1; 1.
 CC PROSITE: PS00262; G_PROTEIN_RECEPTOR_FL_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; 3D-structure.

```

modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
.....
DR EMBL: U09842; AAA18784.1; -.
DR EMBL: AF012103; AAB65428.1; -.
DR HSP: P34936; 1DD.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCRRHODOPSH.
DR PROSITE: PS00237; G_PROTEIN_REC_P1_1; 1.
DR PROSITE: PS50262; G_PROTEIN_REC_P1_2; 1.
KW G-protein coupled receptor; transmembrane; glycoprotein.
FT DOMAIN 1 41
FT TRANSMEM 42 63
FT DOMAIN 64 76
FT TRANSMEM 77 98
FT DOMAIN 99 115
FT TRANSMEM 116 136
FT DOMAIN 137 155
FT TRANSMEM 156 177
FT DOMAIN 178 207
FT TRANSMEM 208 227
FT DOMAIN 228 254
FT TRANSMEM 255 274
FT DOMAIN 275 292
FT TRANSMEM 293 317
FT DOMAIN 318 362
FT DISULFID 113 191
FT CARBOHYD 11 11
FT CARBOHYD 26 26
FT CARBOHYD 102 102
FT CARBOHYD 186 186
SQ SEQUENCE 362 AA; 41180 MW; 3E128B95B64349C CRC64;

Query Match      30.3%; Score 44; DB 1; Length 362;
Best Local Similarity 39.3%; Pred. No. 26;
Matches 11; Conservative 3; Mismatches 12; Indels 2; Gaps 1.
QY      3 REKLRVHTQTGDEYFR--FITLLADPTV 28
Db.     185 RNKTTTCYDTADEYLRSYFVSMCTGV 212

```


DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE Hypothetical protein TV1127.
GN TV1127 OR TWG1156690.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS51 / DSM 4239 / JCM 9571;
RX MEDLINE=20570466; PubMed=1121031.
RA Kawashima T., Amano N., Koike H., Makino S.,
RA Kawashima T., Amano N., Koike H., Makino S.,
RA Nunoshiba T., Yamamoto T., Aramaki H., Makino K., Kawamoto T.,
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT seq. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
KW EMBL: AP000995; BAB60269.1; -;
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 116 AA; 13804 MW; 895956ADB2F96801 CRC64;

```

Query Match 34.5% Score 50; DB 17; Length 116;
Best Local Similarity 39.3% Pred. No. 5.7;
Matches 11; Conservative 6; Mismatches 7; Indels 4; Gaps 1;

QY 5 KLVAVTHTTGDE---YFRFETLLADTV 28
      . . . . . | | | | | | | | | |
32 RDLVYRSTGDRRDYDVNFETLLADVL 59

```

RESULT	7
ID	Q55932
OS	056032.
AC	
PRELIMINARY:	
PRT:	454 AA

DR 01-NOV-1996 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Sensory transduction histidine kinase.
GN SLI0798.
OS *Synechocystis* sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; *Synechocystis*.
OX NCBI_Taxid=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.,
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64 k to 92 k of the genome.";
RN DNA Res. 2:153-166(1995).
RP [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Natsu K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.,
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RN DNA Res. 3:109-136(1996)
CC -I- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
DR EMBL: D64005. BAI0697.1. -
DR InterPro: IPR003594; ATP1nd_Atpase.
DR InterPro: IPR004358; Bact_sens_pr_C.
DR InterPro: IPR0000515; BPD_transp.
DR InterPro: IPR003661; His_kinA.
DR InterPro: IPR004359; His_kinA.
DR Pfam: PF02518; HATPase_g1.
DR Pfam: PF00512; signal; 1.
DR PRINTS: PR00344; BCTRLSENSOR.

DR SMART: SM00387; HATPase_c; 1.
DR SMART: SM00388; HSKA_1.
DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
KW Kinase; Phosphorylation; Sensory transduction; Transference;
KW Complete proteome.
SQ SEQUENCE 454 AA; 50465 MW; F0AB55B5D561D1EE CRC64;

Query Match	34.5%	Score 50;	DB 16;	Length 454.
Best Local Similarity	37.5%	Pred. No. 25;		
Matches	9;	Conservative 5;	Mismatches 10;	Indels 0;
			Gaps 0;	
QY	1 YERERKLRVNHQTGGDEYFRFTLL	24		
	: : : :			
Db	134 FNOETWQTHHPDGGDHYRQFTLL	157		

RESULT 8	
018854	
ID 018854	PRELIMINARY:
CC 018854	PRT: 464 AA

DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE C5464.9 protein.
 GN C5464.9.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae.
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilkinson J.;
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode *C.elegans*: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018 (1998).
 DR EMBL: 275533; CAA99821.1; -
 SO SEQUENCE 464 AA; 5410 MW; 07582BD06C1FCC45B CRC64;

Query March	34.5%	Score 50;	DB 5;	Length 464;
Best Local Similarity	42.9%	Pred. No. 25;		
Matches	9;	Conservative	3;	Mismatches 9; Indels 0; Gaps 0;
QY	2	EREKLRVHQTGDEYFREIT	22	
		: : :		
Db	427	EEVYMRLEHONIGASYFSILT	447	

	RESULT	9
	Q8W040	
ID	Q8W040;	PRELIMINARY; PRT; 571 AA.
AC	Q8W040;	
DT	01-MAR-2002 (TREMBLrel . 20, Created)	
DT	01-MAR-2002 (TREMBLrel . 20, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel . 21, Last annotation update)	
DE	Dihydrofolate synthetase /folypolyglutamate synthetase.	
GN	DHFS /FPGS2.	
OS	Arabidopsis thaliana (mouse-ear cress).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;	
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.	
OX	NCBI_TaxID=3702;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21625122; PubMed=11752472;	
RA	Ravanel S., Cherest H., Jabrin S., Grunwald D., Surdin-Kerjan Y.,	
RA	Douce R., Rebellelle F.:	
RT	"Tetrahydrofolate biosynthesis in plants: Molecular and functional	
RT	characterization of dihydrofolate synthetase and three isoforms of	

```

RT folylpolyglutamate synthetase IN Arabidopsis thaliana."
RL Proc. Natl. Acad. Sci. U.S.A. 98:15360-15365(2001).
DR EMBL: AJ250873; CAC08083.2.
DR InterPro: IPR001645; Polylgl-synthetase.
DR InterPro: IPR000713; Mut_1ligase.
DR Pfam: PF01225; Mut_1ligase.
DR PROSITE: PS01011; FOLYLPOLYGLU-SYNT_1; UNKNOWN_1.
SQ SEQUENCE 571 AA; 63354 MW; AFE9A978A1F45573B CRC64;

Query Match 33.8%; Score 49; DB 10; Length 571;
Best Local Similarity 44.0%; Pred. No. 45;
Matches 11; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

OY 1 YEREKLRVTHQTGDEYFRFFITLLA 25
Db 176 YNRLKERTNEIIPMPYFRFLLA 200

RESULT 10
O61080 PRELIMINARY; PRT; 1186 AA.
AC O61080;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Myosin IC heavy chain.
GN MICHC.
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=88016163; PubMed=3477803;
RA Jung G., Korn E.D., Hammer J.A. III;
RT "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
RT and non-myosin-like sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99079990;
RA Wang Z.Y., Wang F., Sellers J.R., Korn E.D., Hammer J.A. III;
RT "Analysis of the regulatory phosphorylation site in Acanthamoeba
RT myosin IC by using site-directed mutagenesis."
RL Proc. Natl. Acad. Sci. U.S.A. 95:15200-15205(1998).
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AF051353; AAC98089.1;
DR HSSP: P08799; 1MND.
DR InterPro: IPR001609; myosin_head.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PRO0193; MYOSINHEAVY.
DR PRINTS: PRO0452; SH3DOMAIN.
DR ProDom: PD000066; SH3; 1.
DR ProDom: PD000035; myosin_head; 1.
DR SMART: SM00242; MSC; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 1186 AA; 129459 MW; E37AD4A685803A6 CRC64;

Query Match 33.4%; Score 48.5; DB 5; Length 1186;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 13; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

OY 1 YEREKLRVTHQTGDEYFR-FITLLA 25
Db 181 YLLEKSRVYQTNGERNFHFYQLA 206

RESULT 11
O9TSS2 PRELIMINARY; PRT; 83 AA.
ID O9TSS2
```

```

AC O9TSS2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BOLA-DRB3 protein (Fragment).
GN BOLA-DRB3.
OS Bos indicus (zebu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9915;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=BRAMAN; TISSUE=Blood;
RX MEDLINE=99371929; PubMed=10442982;
RA Maillard J.C., Renard C., Chardon P., Chantal I., Bensaïd A.;
RT "Characterization of 18 new BOLA-DRB3 alleles."
RL Anim. Genet. 30:200-204(1999).
DR EMBL: Z82028; CAB52180.1;
DR InterPro: IPR000353; MHC_II_beta.
DR Pfam: PF00969; MHC_II_beta; 1.
DR ProDom: PD000328; MHC_II_beta; 1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 83
SQ SEQUENCE 83 AA; 9975 MW; 5427132A5BA09E03 CRC64;

Query Match 33.1%; Score 48; DB 6; Length 83;
Best Local Similarity 52.6%; Pred. No. 8.1;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 2 EREKLRVTHQTGDEYFRF 20
Db 14 ERVRLDRHFTNGEYFRF 32

RESULT 12
Q84576 PRELIMINARY; PRT; 174 AA.
ID Q84576;
AC Q84576;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE A259L protein.
GN A259L.
OS Paramyxium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=10506;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=96187795; PubMed=8614977;
RA Lu Z., Li Y., Que Q., Kutish G.F., Rock D.L., Van Etten J.L.;
RT "Analysis of 94 kb of the chlorella virus PBCV-1 330-kb genome: map
RT positions 88 to 182."
RL Virology 216:102-123(1996).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=20013326; PubMed=10544099;
RA Kaiser A., Volkmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,
RA Lisek A.D., Nickerson K.W., Van Etten J.L.;
RT "Chlorella virus PBCV-1 encodes a functional homospiridine
RT synthase."
RL Virology 263:254-262(1999).
[3]
RN SEQUENCE FROM N.A.
RX MEDLINE=20478034; PubMed=11021991;
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus
RT PBCV-1."
RL Virology 276:27-36(2000).
[4]
RN SEQUENCE FROM N.A.
RP Van Etten J.L.;
```

RESULT 14	
Q960V7	
ID Q960V7	PRELIMINARY;
AC Q960V7;	PRT; 362 AA

RC STRAIN-BERKELEY;
RX MEDLINE-20196006; Pubmed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wotman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J., H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abul J.F., Abaygani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A.V., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhadrati D., Bolshakov S.,
RA Borkova D., Bolchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadenau E., Center A., Chandra I.,
RA Cherry J.M., Cusley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Foulser K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbalt W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegami C.,
RA Jalali M., Kalush F., Kapten G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Schelker F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rudin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003778; AAF57158.1; -.
 DR FlyBase; FBgn0039851; CG12063.
 DR InterPro; IPR001507; Endoglin/CD105.
 DR InterPro; IPR003014; PAN.
 DR InterPro; IPR003609; Pan_app.
 DR Pfam; PF000024; PAN; 3.
 DR SMART; SM00473; PAN_AP; 3.
 DR SMART; SM00241; ZP; 1.
 SQ SEQUENCE 758 AA; 81720 MW; A7823019671861E3 CRC64;

Query Match 33.1%; Score 48; DB 5; Length 758;
 Best Local Similarity 60.0%; Pred. No. 87;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EREKLRVTHQTGDE 16
 DB 663 ERQKRDVSHQAGDE 677

Search completed: January 29, 2003, 06:50:02
 Job time : 63.8947 secs